BEYOND SUB-GAUSSIAN DETECTOR SCORES: ROBUST WEIGHTED PROFILE-LOSS CHANGE POINT DETECTION FOR HUMAN-LLM TEXT SEGMENTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Mixed human-LLM documents require locating authorship transitions from detector scores whose reliability varies across text units. Existing weighted mean contrasts are vulnerable to extreme scores, while directly replacing means with robust centers obscures how a misplaced boundary changes the population objective. We propose Robust Weighted Profile-Loss Change Point Detection (RWCP), which combines capped reliability weights, Huber profile gains, and narrowest-over-threshold search in reliability coordinates. Our key analysis expresses the population gap between a true and a displaced split as a merge cost, avoiding a closed-form solution for the nonlinear center of a mixed segment. Under explicit curvature, spacing, and dependence conditions, core RWCP recovers the number of changes and localizes their boundaries; its quadratic-loss limit recovers squared weighted CUSUM. We also study RWCP-R, a separately evaluated decoder that shares source centers across nonadjacent passages. Across five retrospective cached-score benchmark families, core RWCP reduces family-macro WindowDiff by 17.6% relative to weighted change-point detection, and RWCP-R lowers it further. Boundary recovery improves most clearly for isolated changes, while both fixed configurations miss changes in collaborative and densely alternating text.

1 INTRODUCTION

Mixed human–LLM writing calls for locating authorship changes rather than labeling an entire document. Human passages can alternate with generated continuations or model-assisted edits, so the output is a sequence of source-homogeneous segments and boundaries. Mapping each sentence or short unit X_i to a detector score $Y_i = \phi(X_i)$ makes this an ordered change-point problem. WCP uses the unequal reliability of these scores (Li et al., 2026), but weighting alone does not control extreme observations or context-induced dependence.

Rare tokens, code, equations, paraphrases, and short units can yield extreme scores, while underestimated variance proxies can give a single unit excessive weight. Huber scores bound their influence, but substituting Huber centers into mean-based CUSUM does not preserve its population decomposition: an incorrect split mixes two source locations, whose optimum is generally not their weighted arithmetic mean. The analysis must compare fitted segments without assuming a formula for that mixed center.

The *profile gain* is the reduction in optimized Huber loss obtained by splitting an interval. With one change, its population difference between the true and a displaced boundary equals a merge cost; local risk curvature makes this gap proportional to misplaced reliability mass and squared jump size. Capped weights control leverage, bounded Huber scores support dependence-aware concentration without sub-Gaussian raw scores, and reliability-coordinate intervals isolate changes by information mass. Guarded recursion converts these properties into count and localization guarantees.

Core RWCP combines robust weighted profile gains with weighted-coordinate narrowest-over-threshold search; in the unconstrained quadratic limit, its gain is exactly squared weighted CUSUM. We prove finite-sample exact count recovery and weighted localization under explicit curvature, central-mass, spacing, and dependence conditions, including the logarithmic cost of geometric

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

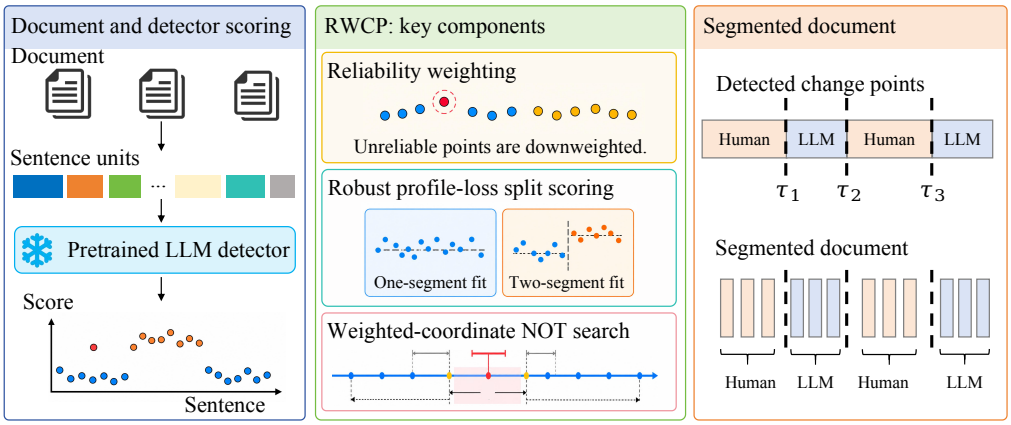


Figure 1: Core RWCP: capped reliability weights, Huber profile gains, and weighted-coordinate narrowest-over-threshold search. The recovery theorem covers this procedure. RWCP-R separately decodes recurrent source states using (6).

mixing. The merge-cost analysis distinguishes this result from a direct robust-center substitution in an existing contrast.

Recurring human and LLM passages offer a second opportunity: nonadjacent passages can share a source state. We study RWCP-R, a distinct decoder with shared centers and a fitted residual scale. We now evaluate an analytic-threshold implementation of core RWCP on the same cached scores as WCP before assessing RWCP-R separately. Core RWCP reduces window error most clearly on controlled single-change texts, but its fixed threshold often misses changes in real collaborative and dense-change documents; RWCP-R improves average window error further without removing this count-recovery limitation. The theorem covers core RWCP under its stated assumptions, not the empirical RWCP-R decoder or the claim that either fixed configuration satisfies those assumptions on detector scores.

Huber profile costs are established in penalized segmentation (**?**), narrowest-over-threshold search supplies interval selection (Baranowski et al., 2019), and WCP supplies the reliability-weighted detector-score setting (Li et al., 2026). Our technical step is to analyze the weighted *difference of optimized profile gains* through a merge-cost margin and carry it through dependence-aware control, reliability-coordinate isolation, and guarded recursion. Related work, full protocols, and proofs appear in Appendices A–C.

2 Robust Weighted Profile Segmentation

We first define a robust gain for testing whether one location or two better describe an interval. We then use that gain in reliability-coordinate interval search; a separate recurrent objective shares centers across passages assigned to the same source state.

2.1 Robust Locations, Reliability, and Profile Gains

Let $0 = \tau_0 < \tau_1 < \cdots < \tau_K < \tau_{K+1} = N$ denote unknown boundaries. Observation i has a deterministic reliability weight $w_i > 0$, scale $s_i = w_i^{-1/2}$, and segment index $z(i)$. For cutoff $c > 0$, the Huber loss and score are

$$\rho_c(u) = \begin{cases} u^2/2, & |u| \le c, \\ c|u| - c^2/2, & |u| > c, \end{cases} \qquad \psi_c(u) = \operatorname{sign}(u)\min\{|u|, c\}.$$

Definition 2.1 (Piecewise common Huber location)**.** There are locations $\theta_1, \ldots, \theta_{K+1}$ in a compact interval Θ such that, whenever $\tau_{j-1} < i \le \tau_j$, θ_j minimizes $Q_i(\theta) = \mathbb{E}\rho_c\{\sqrt{w_i}(Y_i - \theta)\}$. For interior unique minimizers this is equivalent to

$$\mathbb{E}[\sqrt{w_i}\,\psi_c\{\sqrt{w_i}(Y_i - \theta_j)\}] = 0. \tag{1}$$

Adjacent jump sizes are $\kappa_j = |\theta_{j+1} - \theta_j|$.

The location is defined by the detector-score risk rather than by an ordinary expectation. A symmetric location family $Y_i = \theta_{z(i)} + \sigma_i \xi_i$ is sufficient, but asymmetric distributions are allowed whenever the centering condition in (1) holds. Given a variance proxy σ_i^2, the capped oracle weight is

$$w_i^\circ = \Pi_{[w_{\min}, w_{\max}]}(\sigma_i^{-2}), \qquad \Pi_{[a,b]}(u) = \min\{\max(u, a), b\}. \tag{2}$$

Practical weights replace σ_i^2 by an externally estimated proxy and add a small denominator ridge. Capping in (2) prevents a single underestimated variance from receiving unbounded leverage. Write $S^w_{a:b} = \sum_{i=a}^b w_i$ and $d_w(u, v) = \sum_{\min(u,v)<i\le\max(u,v)} w_i$.

For an interval A, define $\widehat{\mathcal{L}}_A(\theta) = \sum_{i\in A} \rho_c\{\sqrt{w_i}(Y_i - \theta)\}$ and let $\widehat{\theta}_A$ minimize it over Θ. For $s \le b < e$, RWCP uses the profile improvement

$$\widehat{\Gamma}_{s,e}(b) = 2\Big\{\widehat{\mathcal{L}}_{s:e}(\widehat{\theta}_{s:e}) - \widehat{\mathcal{L}}_{s:b}(\widehat{\theta}_{s:b}) - \widehat{\mathcal{L}}_{b+1:e}(\widehat{\theta}_{b+1:e})\Big\}. \tag{3}$$

The gain in (3) is nonnegative because the one-location model is nested in the split model. Its population analogue makes a displaced split pay for merging two distinct source locations, the mechanism quantified in Theorem 3.8.

Proposition 2.2 (Quadratic-loss reduction)**.** *If ρ_c is replaced by $\rho_\infty(u) = u^2/2$ and the three fits are unconstrained, or their weighted means all lie in Θ, then*

$$\widehat{\Gamma}_{s,e}(b) = \frac{S^w_{s:b} S^w_{b+1:e}}{S^w_{s:e}} \left(\bar{Y}^w_{s:b} - \bar{Y}^w_{b+1:e}\right)^2 = \{W^Y_{s,e}(b)\}^2, \tag{4}$$

where $\bar{Y}^w_{a:b} = \sum_{i=a}^b w_i Y_i / S^w_{a:b}$ and $W^Y_{s,e}(b)$ is the weighted CUSUM.

Active parameter constraints introduce projection terms (Appendix C.3.2). Thus RWCP retains the familiar WCP contrast when its loss is quadratic, while the finite-cutoff analysis uses bounded Huber scores in place of a raw-score tail condition.

2.2 Weighted-coordinate search and recurrent states

The reliability mass used by the profile gain also defines the search geometry. RWCP samples intervals uniformly in the cumulative coordinate $W(t) = \sum_{i\le t} w_i$, rather than uniformly in sentence index. Within $I = [s, e]$, it scans splits whose two sides each have mass at least m_N, computes $T(I) = \max_b \widehat{\Gamma}_{s,e}(b)$, and retains intervals with $T(I) > r_N$. Among them it selects the interval with the least reliability mass, reports its maximizing split $b^\star$, and recurses. This is the narrowest-over-threshold principle (Baranowski et al., 2019) in the information geometry supplied by the detector.

After accepting a boundary, a guard of mass g_N prevents duplicate discoveries while preserving well-separated neighboring changes. For a current search range $[a, b]$, define the nearest mass crossings

$$\begin{aligned} t_L &= \max\{t \in [a, b^\star] : S^w_{t:b^\star} \ge g_N\}, & \ell_g &= t_L - 1, \\ t_R &= \min\{t \in [b^\star + 1, b] : S^w_{b^\star+1:t} \ge g_N\}, & r_g &= t_R + 1. \end{aligned} \tag{5}$$

If the left or right crossing does not exist, set $\ell_g = a - 1$ or $r_g = b + 1$, respectively. Recursion continues on the nonempty ranges $[a, \ell_g]$ and $[r_g, b]$. Each removed side has mass below $g_N + w_{\max}$. The theorem uses the analytic threshold $r_N = C_r \Lambda_N(\delta)$. In practice, a block-multiplier alternative forms clipped influence residuals from a conservative pilot, multiplies overlapping residual blocks by Gaussian weights, and takes the conditional quantile of the maximum over the same fixed candidate collection. This calibration is not part of the recovery theorem.

When source states recur, nonadjacent passages can inform a common center. RWCP-R uses this structure through a separate recurrent-state objective; it replaces the core interval search and does not invoke its threshold or guard. With robustly standardized scores $\widetilde{Y}_i$, compressed capped weights $\widetilde{w}_i$, S source states, centers μ, and scale $\sigma \ge \sigma_{\min}$, it seeks a local minimum of

$$\mathcal{J}(z, \mu, \sigma) = \sum_{i=1}^N \rho_c\left(\frac{\sqrt{\widetilde{w}_i}(\widetilde{Y}_i - \mu_{z_i})}{\sigma}\right) + N \log \sigma + \lambda \sum_{i=2}^N \mathbf{1}\{z_i \ne z_{i-1}\}. \tag{6}$$

For fixed centers and scale, Viterbi decoding solves the state update in $O(NS^2)$; for fixed states, weighted Huber fits update the centers and a monotone score equation updates the scale. Exact conditional updates cannot increase (6); the joint problem is nonconvex and the finite-iteration solver has no global-optimality guarantee. Here σ is common to all states. With residuals $r_i = \sqrt{\widetilde{w}_i}(\widetilde{Y}_i - \mu_{z_i})/\sigma$, an interior scale update solves $\sum_i \psi_c(r_i) r_i = N$; otherwise the constrained optimum is $\sigma_{\min}$. We declare $S = 3$ for CoAuthor and $S = 2$ for the controlled tasks. Boundaries are state transitions; S specifies source classes, not their number of occurrences. Appendix B.3 gives the finite-iteration algorithm. Section 3 analyzes core RWCP; Section 4 evaluates the recurrent decoder as a separate empirical extension.

3 Theoretical Guarantees

The recovery argument follows the algorithm's four steps: profile-loss geometry supplies a population margin; bounded Huber scores and empirical curvature preserve it in the observed data; reliability-coordinate intervals isolate each change; and a guard prevents duplicate detections during recursion. We state the assumptions and main result here, with verification results, lower bounds, constants, and proofs in Appendix C.

Assumption 3.1 (Robust locations)**.** Definition 2.1 and (1) hold with finite risks. There is a fixed radius $r_0 > 0$ such that $\operatorname{dist}(\theta_j, \partial\Theta) \geq r_0$ for every true center.

Assumption 3.2 (Risk curvature)**.** There are fixed $0 < m_0 \leq M_0 < \infty$ such that, for every i in segment j and every $|\theta - \theta_j| \leq r_0$,

$$\frac{m_0 w_i}{2}(\theta - \theta_j)^2 \leq Q_i(\theta) - Q_i(\theta_j) \leq \frac{M_0 w_i}{2}(\theta - \theta_j)^2. \tag{7}$$

Assumption 3.3 (Reliability)**.** The deterministic weights satisfy $0 < w_{\min} \leq w_i \leq w_{\max} < \infty$, and the minimum segment mass $\Delta_w = \min_j S^w_{\tau_{j-1}+1:\tau_j}$ obeys $w_{\max} \leq \Delta_w/24$.

Assumption 3.4 (Dependence)**.** Every centered bounded transform $V_i = f_i(Y_i)$ used for scores or central-mass indicators obeys, on every interval A,

$$\mathbb{P}\left(\left|\sum_{i\in A} V_i\right| > C_{\mathrm{dep}}\|V\|_\infty\{\sqrt{|A|x} + q_N x\}\right) \leq 2e^{-x}, \qquad x \geq 1.$$

For a two-sided extension, absolute regularity is measured by

$$\beta(k) = \sup_t \mathbb{E}\left[\sup_{B\in\mathcal{F}^\infty_{t+k}} \left|\mathbb{P}(B \mid \mathcal{F}^t_{-\infty}) - \mathbb{P}(B)\right|\right]. \tag{8}$$

Independent and fixed-order dependent observations permit $q_N = O(1)$; the conservative geometric-mixing specialization $\beta(k) \lesssim e^{-ck}$ uses $q_N = O\{\log^2(eN)\}$.

Assumption 3.5 (Central mass)**.** There is a fixed $p_0 \in (0, 1]$ such that every observation i in segment j and every $|u| \leq r_0$ satisfy $\mathbb{P}\big(\big|\sqrt{w_i}(Y_i - \theta_j - u)\big| \leq c/2\big) \geq p_0$. For $K \geq 1$, we impose the local-jump restriction $\kappa_{\max} \leq p_0 r_0/16$; for $K = 0$, set $\kappa_{\max} = 0$.

Let $\mathcal{A}_N$ contain all nonempty index intervals and let $L_N = 1 + N^2 + N^3$ bound their number together with all split triples. Write $x_N(\delta) = \log(64L_N/\delta)$ and define

$$\Lambda_N(\delta) = q_N^2 \log\left(\frac{64L_N}{\delta}\right).$$

Set $m_{\mathrm{curv}} = C_{\mathrm{curv}}\Lambda_N(\delta)$.

Assumption 3.6 (Spacing and signal)**.** For $K \geq 1$, set $\kappa_{\min} = \min_j \kappa_j > 0$ and $\kappa_{\max} = \max_j \kappa_j$. For constants $c_{m,1}, C_g, C_{\mathrm{snr}} > 0$ and $0 < c_{m,2} \leq 1/12$, the admissible side mass m_N, guard mass g_N, and signal obey

$$m_{\mathrm{curv}} \leq m_N \leq c_{m,1}\frac{\Lambda_N(\delta)}{\kappa^2_{\max}}, \qquad C_g\frac{\Lambda_N(\delta)}{\kappa^2_{\min}} + w_{\max} \leq g_N \leq c_{m,2}\Delta_w,$$
$$m_N \leq \frac{\Delta_w}{12}, \qquad \kappa^2_{\min}\Delta_w \geq C_{\mathrm{snr}}\Lambda_N(\delta).$$

For adjacent pure intervals A, B, let $D^\star(A, B)$ be the excess population loss caused by fitting their union with one location instead of two.

Lemma 3.7 (Mixed-segment excess risk)**.** *If A and B have locations θ_A, θ_B in the common curvature region, then*

$$D^\star(A, B) \geq \frac{m_0}{2} \frac{S_A^w S_B^w}{S_A^w + S_B^w} (\theta_A - \theta_B)^2, \tag{9}$$

$$D^\star(A, B) \leq \frac{M_0}{2} \frac{S_A^w S_B^w}{S_A^w + S_B^w} (\theta_A - \theta_B)^2. \tag{10}$$

Theorem 3.8 (Single-change population margin)**.** *Suppose $[s, e]$ contains one change τ of size κ. The population counterpart of* (3) *is maximized at τ. For $b < \tau$,*

$$\Gamma^\star_{s,e}(\tau) - \Gamma^\star_{s,e}(b) = 2D^\star([b+1, \tau], [\tau+1, e]) \geq m_0 \kappa^2 \frac{S^w_{b+1:\tau} S^w_{\tau+1:e}}{S^w_{b+1:\tau} + S^w_{\tau+1:e}}, \tag{11}$$

with the symmetric bound for $b > \tau$. When the misclassified mass is no larger than the opposite pure-side mass, the gap is at least $m_0 \kappa^2 d_w(b, \tau)/2$.

The identity in (11) converts a difficult mixed-center calculation into an excess-risk comparison. Bounded Huber scores keep empirical same-location merge gains at $O\{\Lambda_N(\delta)\}$, while different-location merges retain the curvature term in Lemma 3.7.

Lemma 3.9 (Uniform empirical merge costs)**.** *For adjacent intervals A, B with $S_A^w, S_B^w \geq m_{\mathrm{curv}}$, define*

$$\widehat{D}(A, B) = \inf_\theta \{\widehat{\mathcal{L}}_A(\theta) + \widehat{\mathcal{L}}_B(\theta)\} - \inf_\theta \widehat{\mathcal{L}}_A(\theta) - \inf_\theta \widehat{\mathcal{L}}_B(\theta).$$

On an event of probability at least $1 - \delta/2$, uniformly over the master collection, intervals from the same true segment with masses at least m_{curv} satisfy

$$0 \leq \widehat{D}(A, B) \leq C_0 \Lambda_N(\delta). \tag{12}$$

If instead A, B lie in adjacent segments separated by κ, then

$$\widehat{D}(A, B) \geq c_0 \kappa^2 \frac{S_A^w S_B^w}{S_A^w + S_B^w} - C_1 \Lambda_N(\delta), \tag{13}$$

$$\widehat{D}(A, B) \leq C_2 \kappa^2 \frac{S_A^w S_B^w}{S_A^w + S_B^w} + C_3 \Lambda_N(\delta). \tag{14}$$

Thus the same gain separates homogeneous intervals from intervals containing a sufficiently strong change: bounded scores control the null merge cost, and empirical curvature preserves the κ^2-weighted alternative up to a uniform fluctuation.

Let $W_N = W(N)$ denote the total reliability mass. Weighted-coordinate sampling supplies the algorithmic counterpart of the population margin.

Lemma 3.10 (Weighted-coordinate isolation)**.** *Under Assumption 3.3, suppose $\max_i w_i \leq \Delta_w/24$. If the number of sampled intervals satisfies*

$$M \geq C_I \left(\frac{W_N}{\Delta_w}\right)^2 \log\left\{\frac{4(K \vee 1)}{\delta}\right\}, \tag{15}$$

then, with probability at least $1 - \delta/4$, every change τ_j is the only change in some sampled interval $[s, e]$ for which

$$\frac{\Delta_w}{6} \leq S^w_{s:\tau_j},\ S^w_{\tau_j+1:e} \leq \frac{\Delta_w}{3}, \qquad S^w_{s:e} \leq \frac{2\Delta_w}{3}. \tag{16}$$

Theorem 3.11 (Finite-sample count recovery and localization)**.** *Under Assumptions 3.1–3.6, sample M weighted-coordinate intervals with M satisfying* (15)*, and choose $r_N = C_r \Lambda_N(\delta)$ with the constants specified in Appendix C.2.3. Then, with probability at least $1 - \delta$, core RWCP returns exactly K changes and, after ordering,*

$$d_w(\widehat{\tau}_j, \tau_j) \leq C_{\mathrm{loc}} \frac{\Lambda_N(\delta)}{\kappa_j^2}, \qquad j \in [K]. \tag{17}$$

For $K = 0$, Assumptions 3.1–3.5, $m_N \geq m_{\mathrm{curv}}$, and $r_N > 2C_0\Lambda_N$ suffice for no false boundary; the jump-dependent spacing and interval-count conditions are unnecessary.

The theorem describes locally separated changes; feasible side mass requires $C_{\mathrm{curv}}\kappa_{\max}^2 \le c_{m,1}$, and g_N, M use signal/spacing bounds rather than fully adaptive tuning. In the proof, uniform merge-cost bounds suppress homogeneous intervals, while weighted isolation supplies a short over-threshold interval for each change. The narrowest such interval contains one change; the profile margin localizes it, and the guard removes it without losing neighboring isolating intervals. Induction yields exact K and simultaneous localization.

Corollary 3.12 (Estimated-weight transfer)**.** *Suppose a weight estimator obeys $c_w w_i^\circ \le \widehat{w}_i \le C_w w_i^\circ$ uniformly with probability at least $1-\delta_w$, and conditional score assumptions hold uniformly on this event. Then RWCP computed with $\widehat{w}_i$ satisfies, with probability at least $1-\delta-\delta_w$,*

$$\widehat{K} = K, \qquad d_{w^\circ}(\widehat{\tau}_j, \tau_j) \le c_w^{-1} d_{\widehat{w}}(\widehat{\tau}_j, \tau_j) \le C\frac{\Lambda_N(\delta)}{\kappa_j^2}.$$

The spacing, no-dominant-atom, and signal conditions change only by constants depending on (c_w, C_w).

Independent and fixed-order dependent scores attain weighted localization order $O\{\log(N/\delta)/\kappa_j^2\}$; geometric mixing incurs the explicit q_N^2 factor. External proxies or sample splitting can justify Corollary 3.12; target-score-derived weights require separate joint analysis.

Under independent Gaussian scores and inactive caps, the two-point lower bound matches the $\log N/\kappa^2$ order (Appendix C.3.1); it does not address dependence or active caps.

The weighted metric translates to text units: bounded weights make $d_w(u, v)$ comparable to the number of misplaced sentences, while $\sigma_i^2 \asymp n_i^{-1}$ with inactive caps makes it comparable to misplaced tokens. Formal metric corollaries and the lower-bound argument appear in Appendix C.3.1.

4 Experiments

4.1 Protocol and overall comparison

We compare an analytic-threshold implementation of core RWCP with WCP, then evaluate whether the separate recurrent decoder improves window segmentation and how both methods handle dense changes. Five cached-score families cover single changes in WikiQA, News, and Story; real CoAuthor sessions; multiple-change Story documents; text attacks; and varying LLM-written proportions. The suite has 25 settings and 2,690 instances, including all WikiQA records under the common singleton convention (Appendix B.1). Comparators are TextTiling, direct and sentence-level prediction, local voting, VCP, and WCP where outputs are available (Hearst, 1997; Zhang et al., 2024; Li et al., 2026). Core RWCP, RWCP-R, and WCP share cached scores and reliability proxies on nontrivial inputs; single-change experiments use fine-tuned scores and the other families use normalized scores. Core RWCP uses one fixed analytic-threshold configuration, specified in Appendix B.2; its profile search is the subject of Theorem 3.11, although these numerical settings are not certified to satisfy the theorem's unknown constants and assumptions. RWCP-R uses a separately frozen configuration: Huber cutoff $c = 1.5$, transition penalty $\lambda = 4$, weight exponent $\eta = 0.5$, scale floor 0.25, and six deterministic starts. Appendix B.4 documents its selection on benchmarks that informed earlier development, making the comparisons retrospective.

We use standard raw-boundary WindowDiff (WD, lower is better) to measure window-level disagreement and raw count error $\mathrm{CE}_d = K_d - \widehat{K}_d$ to show its direction. Positive CE indicates missed changes and negative CE indicates excess changes; count MAE, matched-boundary F1, and a no-boundary reference complete the recovery assessment in Appendices B.2 and B.7.2. The inherited final-label window error $\mathrm{WD}_{\mathrm{sup}}$ is reported separately because it can exceed one and measures a different output.

Core RWCP lowers family-macro WD from 0.4070 for WCP to 0.3355 (17.6% relative) and improves WD in four of five families, but these window-level gains do not imply reliable boundary recovery. Its family-macro count MAE rises from 2.808 to 3.229 and matched-boundary F1 falls from 0.287 to 0.129. Single-change texts are the family-level exception: count MAE improves from 1.637 to 0.393 and F1 from 0.435 to 0.487, although F1 does not improve in every domain. On CoAuthor, multiple-change Story, text attacks, and LLM-proportion documents, core RWCP reports no boundary in 82.4%,

Table 1: Raw-boundary results averaged over documents, then equally over settings within each family (300, 290, 1,000, 600, and 500 instances). Core RWCP uses the fixed analytic threshold; the historical RWCP-analytic/block variants in Appendix B.7 use additional practical refinements. Bold marks the lowest unrounded WD, not significance; CE is signed count bias. Dashes denote unavailable complete-family results.

Method	Single change		CoAuthor		Multiple changes		Text attacks		LLM proportion	
	WD ↓	CE	WD ↓	CE	WD ↓	CE	WD ↓	CE	WD ↓	CE
TextTiling	0.6449	-2.487	0.5445	7.314	0.6251	-4.585	0.6421	-2.478	0.7611	-3.730
LLMPred	0.4446	-0.807	0.4996	8.303	0.4967	-0.141	0.4508	-1.013	0.4074	-1.356
SenPred	0.6297	-6.717	0.5485	-4.155	0.9032	-23.498	0.6791	-7.390	0.9511	-14.500
Voting	0.2725	-0.903	0.5487	3.286	0.4797	-4.626	0.3514	-1.338	0.4939	-3.164
VCP	0.3086	-0.603	0.4913	9.369	0.4613	-0.579	–	–	–	–
WCP	0.3961	-1.403	0.4878	7.717	0.4363	0.772	0.3609	-0.358	0.3538	-0.804
Core RWCP	0.2315	0.220	0.4897	10.152	0.4108	3.716	0.3369	0.940	**0.2085**	0.920
RWCP-R	**0.1895**	0.280	**0.4869**	10.166	**0.3993**	3.181	**0.2676**	0.475	0.2119	0.600

93.1%, 94.7%, and 92.8% of instances, respectively. Appendix Table 6 reports these diagnostics and conditional localization errors; no CoAuthor document has the correct positive change count under this fixed configuration. These core comparisons are descriptive; the paired intervals below concern RWCP-R.

RWCP-R has lower mean WD than WCP in all five families and lower WD than the fixed core implementation in four (Table 1). With equal family weights, WD falls from 0.4070 for WCP to 0.3110, a 23.6% relative reduction; legacy $\mathrm{WD}_{\mathrm{sup}}$ falls from 0.3249 to 0.2990 (8.0%). The frozen RWCP-R configuration improves standard WD in 23 of 25 settings against WCP and in 20 against every available non-RWCP baseline. Nineteen pointwise paired 95% intervals for the WD difference against WCP lie below zero, one lies above zero, and five include or touch zero. The joint source-component bootstrap gives a family-macro difference of -0.0959 with interval $[-0.1074, -0.0844]$ (Appendix Table 19); these exploratory intervals condition on the saved configurations and available source linkage.

Window-level agreement and recovery of individual changes answer different questions, so we interpret the two together. RWCP-R improves matched-boundary F1 over WCP on single-change and text-attack families, from 0.435 to 0.544 and from 0.272 to 0.300, respectively. On CoAuthor and dense-change Story, the fitted transition penalty yields fewer boundaries and lower F1 despite the WD gains; Appendix Table 16 reports the corresponding count MAE, F1, and no-boundary rates. This distinction motivates the density analysis below and identifies the regimes where the recurrent model's window-level advantage translates most directly into localized boundaries.

4.2 Controlled localization and text perturbations

The single-change and attack settings test whether the same configuration transfers across three domains and two text perturbations (Table 2). RWCP-R attains the lowest available standard WD in all three unperturbed domains, all three decoherence settings, and two of three paraphrasing settings; Voting is better on paraphrased News. All nine paired WD intervals against WCP exclude zero in favor of RWCP-R. Count bias adds a useful qualification: RWCP-R is closer to zero than WCP on unperturbed News and Story, while it is more conservative on WikiQA and the decoherence settings. Together with the higher family-level boundary F1, these results identify single changes and attacked text as the clearest empirical successes of the recurrent decoder. They compare complete frozen pipelines; the component analysis below examines individual design choices at that configuration. The GPT-5-mini single-change aggregates remain a separate supplemental panel (Table 10), outside the primary means and paired analysis.

4.3 Change density and real collaborative text

The multiple-change settings probe the transition from isolated changes to short alternating passages (Table 3). Relative to WCP, standard WD improves for Claude at $K \in \{1, 2, 3\}$ and for GPT-5-mini at every evaluated K. The advantage narrows at high density: at $K = 8$, the WD difference is $+0.018$

Table 2: Controlled Claude Haiku 4.5 results: 100 records per domain and condition under the common singleton convention. Bold marks the lowest unrounded WD; CE is signed count bias. Dashes indicate unavailable outputs. Legacy metrics appear in Tables 9 and 13.

Method	WikiQA		News		Story	
	WD ↓	CE	WD ↓	CE	WD ↓	CE
Single change						
TextTiling	0.427	-0.09	0.796	-4.18	0.712	-3.19
LLMPred	0.428	0.41	0.420	-0.81	0.486	-2.02
PaLD	0.431	-1.28	–	–	–	–
SenPred	0.257	-0.90	0.707	-6.74	0.924	-12.51
Voting	0.286	0.00	0.107	-0.32	0.424	-2.39
VCP	0.267	0.20	0.265	-0.99	0.394	-1.02
WCP	0.284	0.07	0.370	-1.73	0.534	-2.55
RWCP-R	**0.223**	0.47	**0.102**	0.03	**0.243**	0.34
Decoherence						
TextTiling	0.427	-0.09	0.797	-4.18	0.708	-3.17
LLMPred	0.416	0.37	0.437	-1.22	0.542	-2.96
SenPred	0.362	-1.30	0.885	-10.58	0.953	-12.89
Voting	0.359	-0.03	0.363	-1.75	0.639	-3.82
WCP	0.340	0.44	0.407	-0.59	0.398	-0.45
RWCP-R	**0.291**	0.65	**0.335**	0.60	**0.335**	0.59
Paraphrasing						
TextTiling	0.431	-0.06	0.785	-4.16	0.704	-3.21
LLMPred	0.429	0.30	0.395	-0.44	0.486	-2.13
SenPred	0.273	-0.96	0.696	-6.53	0.905	-12.08
Voting	0.267	0.04	**0.091**	-0.24	0.389	-2.23
WCP	0.266	0.25	0.326	-1.05	0.428	-0.75
RWCP-R	**0.189**	0.37	0.146	0.12	**0.310**	0.52

Table 3: Story change-density results: 100 documents per generator and K. True K is used only for evaluation. WD is standard raw-boundary WindowDiff; CE is signed count bias. Bold compares WCP and RWCP-R within each generator.

Generator	K	WCP		RWCP-R	
		WD	CE	WD	CE
Claude Haiku 4.5	1	0.400	-0.52	**0.288**	0.58
	2	0.404	-0.19	**0.374**	1.44
	3	0.425	-0.11	**0.378**	2.10
	5	**0.425**	0.39	0.429	4.26
	8	**0.430**	2.85	0.448	7.21
GPT-5-mini	1	0.434	-0.38	**0.339**	0.79
	2	0.430	0.01	**0.408**	1.51
	3	0.470	0.31	**0.418**	2.42
	5	0.475	1.26	**0.447**	4.28
	8	0.471	4.10	**0.464**	7.22

for Claude and -0.006 for GPT-5-mini. CE exceeds seven in both cases, corresponding to fewer than one detected boundary on average despite eight reference changes. The trend is consistent with the practical difficulty of separating short passages under a fixed transition penalty and shows where the window-level advantage becomes small.

On CoAuthor, WD changes only from 0.4878 to 0.4869, while mean CE increases from 7.717 to 10.166. The no-boundary predictor is close on CoAuthor and the LLM-proportion family: its WD is 0.4893 and 0.2141, compared with 0.4869 and 0.2119 for RWCP-R. Thus the strongest evidence for localized boundary improvement comes from the single-change and perturbation families, whereas these two families emphasize the distinction between WD and exact boundary recovery. Appendix Tables 11, 14, and 16 give the full baseline, proportion, and recovery diagnostics.

4.4 COMPONENT CONTRIBUTIONS

Table 4: Frozen component comparisons, averaged over settings and then families. Fixed scale uses $\sigma = 1$, uniform weights exponent zero, and quadratic loss $c = 10^6$. No variant is retuned. Bold marks the lowest window error; signed CE is not ranked.

Variant	Raw boundaries		Final labels
	WD↓	Mean CE	$\mathrm{WD}_{\mathrm{sup}}\downarrow$
RWCP-analytic	0.3466	2.915	0.3233
RWCP-block	0.3371	3.050	0.3158
RWCP-R	**0.3110**	2.940	**0.2990**
R: fixed scale	0.3170	2.984	0.3038
R: uniform weights	0.3118	2.756	0.3058
R: quadratic approximation	0.3179	2.782	0.3130

Table 4 probes three design choices at the selected recurrent configuration. Fitting the scale and retaining the finite Huber cutoff each improve both window errors relative to their fixed-configuration alternatives. Uniform weights nearly match standard WD: the paired difference for full RWCP-R minus uniform weights is -0.0008, with interval $[-0.0052, 0.0038]$. This contrast makes the empirical role of weighting metric-specific: the full model has lower legacy window error, while uniform weights have lower count bias and stronger boundary matching in Appendix Table 17.

RWCP-R also reduces family-macro standard WD by 10.3% and 7.7% relative to the historical analytic and block-calibrated variants. These are different from the fixed analytic-threshold core evaluation in Table 1. Because those historical variants differ in search, scale, and other settings together, we interpret this comparison as a pipeline result; the fixed component variants provide the more focused evidence about scale, loss, and weighting. Family-wise comparisons, auxiliary metrics, and the post-freeze stability check are reported in Appendix B.7.

5 DISCUSSION AND CONCLUSION

RWCP turns robust change-point detection into a comparison of optimized segment risks. The merge-cost identity supplies a population margin without solving for nonlinear mixed-segment Huber centers, and the quadratic limit recovers weighted CUSUM. With bounded-score concentration, reliability-coordinate isolation, and guarded recursion, this yields exact count recovery and weighted localization for core RWCP under the stated conditions. The new fixed-configuration core evaluation directly measures the corresponding profile search on cached detector scores, but it does not establish that those scores or numerical thresholds meet the theorem's assumptions.

The core evaluation lowers average WindowDiff relative to WCP, with a simultaneous count and boundary-F1 gain only on the single-change family. Its high no-boundary rates on CoAuthor and dense or altered texts demonstrate the limits of a fixed analytic threshold in this retrospective score suite. The recurrent extension shares information across nonadjacent passages and lowers family-macro WindowDiff further, with the clearest boundary-recovery gain on single changes; its count/F1 tradeoff persists in other families. These results motivate count-aware calibration on independent detector streams.

The two contributions have complementary scopes: Theorem 3.11 covers analytic-threshold core RWCP under specified conditions, while RWCP-R evaluates a distinct recurrent objective on retrospective cached scores. The empirical results support window-error reductions, not general exact-count recovery on this benchmark suite; gradual editing, within-sentence mixing, reversed score semantics, long-memory dependence, and target-dependent weight estimation remain outside the present score model.

ETHICS STATEMENT

RWCP is a research tool for provenance segmentation, not automatic evidence of misconduct. Detector errors can vary across language, genre, dialect, and demographic writing style; deployment

therefore requires domain-specific validation, uncertainty reporting, human review, and respect for dataset and model-provider terms.

Reproducibility Statement

The appendices document assumptions, proofs, frozen settings, and per-setting results. The accompanying local records preserve cached inputs and outputs, metric-recomputation scripts, and the complete-record correction; they do not constitute a fresh independent test or establish exact regeneration from model names.

AI Usage Statement

Generative AI assisted with manuscript editing, LaTeX, code, and consistency checks, including the reporting correction. Numerical comparisons are derived from saved predictions and explicit evaluation conventions. The authors remain responsible for validating the mathematics, implementation, citations, and empirical claims.

References

Guangsheng Bao, Yanbin Zhao, Zhiyang Teng, Linyi Yang, and Yue Zhang. Fast-DetectGPT: Efficient zero-shot detection of machine-generated text via conditional probability curvature. In *International Conference on Learning Representations*, 2024.

Rafal Baranowski, Yining Chen, and Piotr Fryzlewicz. Narrowest-over-threshold detection of multiple change points and change-point-like features. *Journal of the Royal Statistical Society: Series B*, 81 (3):649–672, 2019.

Haeran Cho and Claudia Kirch. Two-stage data segmentation permitting multiscale change points, heavy tails and dependence. *Annals of the Institute of Statistical Mathematics*, 74(4):653–684, 2022.

Paul Doukhan. *Mixing: Properties and Examples*. Springer, 1994.

Angela Fan, Mike Lewis, and Yann Dauphin. Hierarchical neural story generation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pp. 889–898, 2018. doi: 10.18653/v1/P18-1082.

Piotr Fryzlewicz. Wild binary segmentation for multiple change-point detection. *The Annals of Statistics*, 42(6):2243–2281, 2014.

Sebastian Gehrmann, Hendrik Strobelt, and Alexander M. Rush. GLTR: Statistical detection and visualization of generated text. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 111–116, 2019.

Abhimanyu Hans, Avi Schwarzschild, Valeriia Cherepanova, Hamid Kazemi, Aniruddha Saha, Micah Goldblum, Jonas Geiping, and Tom Goldstein. Spotting LLMs with binoculars: Zero-shot detection of machine-generated text. In *International Conference on Machine Learning*, 2024.

Marti A. Hearst. TextTiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics*, 23(1):33–64, 1997.

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

Peter J. Huber and Elvezio M. Ronchetti. *Robust Statistics*. Wiley, 2 edition, 2009.

Hans R. Künsch. The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3):1217–1241, 1989.

Mina Lee, Percy Liang, and Qian Yang. CoAuthor: Designing a human–AI collaborative writing dataset for exploring language model capabilities. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–19, 2022.

Enzo Lei, Hsiang Hsu, and Ching-Feng Chen. PaLD: Detection of text partially written by large language models. In *International Conference on Learning Representations*, 2025.

Mengchu Li and Yi Yu. Adversarially robust change point detection. In *Advances in Neural Information Processing Systems*, volume 34, pp. 22955–22967, 2021.

Mengchu Li, Jin Zhu, Jinglai Li, and Chengchun Shi. Segmenting human–LLM co-authored text via change point detection. *arXiv preprint arXiv:2605.03723*, 2026.

Florence Merlevède, Magda Peligrad, and Emmanuel Rio. Bernstein inequality and moderate deviations under strong mixing conditions. In *High Dimensional Probability V: The Luminy Volume*, pp. 273–292. Institute of Mathematical Statistics, 2009.

Eric Mitchell, Yoonho Lee, Alexander Khazatsky, Christopher D. Manning, and Chelsea Finn. DetectGPT: Zero-shot machine-generated text detection using probability curvature. In *International Conference on Machine Learning*, pp. 24950–24962, 2023.

Shashi Narayan, Shay B. Cohen, and Mirella Lapata. Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 1797–1807, 2018. doi: 10.18653/v1/D18-1206.

Lev Pevzner and Marti A. Hearst. A critique and improvement of an evaluation metric for text segmentation. *Computational Linguistics*, 28(1):19–36, 2002.

Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. SQuAD: 100,000+ questions for machine comprehension of text. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 2383–2392, 2016. doi: 10.18653/v1/D16-1264.

Eduard Tulchinskii, Kristian Kuznetsov, Laida Kushnareva, Daniil Cherniavskii, Sergey Nikolenko, Evgeny Burnaev, Serguei Barannikov, and Irina Piontkovskaya. Intrinsic dimension estimation for robust detection of AI-generated texts. In *Advances in Neural Information Processing Systems*, volume 36, pp. 39257–39276, 2023.

Daren Wang, Yi Yu, and Alessandro Rinaldo. Univariate mean change point detection: Penalization, CUSUM and optimality. *Electronic Journal of Statistics*, 14(1):1917–1961, 2020.

Yi Yu. A review on minimax rates in change point detection and localisation. *arXiv preprint arXiv:2011.01857*, 2020.

Zijie Zeng, Shuyu Liu, Lei Sha, Zhuang Li, Kaixuan Yang, Sheng Liu, Dragan Gašević, and Guanliang Chen. Detecting AI-generated sentences in human–AI collaborative hybrid texts: Challenges, strategies, and insights. In *International Joint Conference on Artificial Intelligence*, 2024.

Zhenyu Zhang, Wen Qin, and Bryan A. Plummer. Machine-generated text localization. In *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 8357–8371, 2024.

A Related Work

This section positions RWCP at the intersection of machine-text detection, structured authorship localization, robust change point analysis, and dependence-aware calibration. The comparison clarifies which ingredients are inherited from detector scoring and weighted segmentation, and which are introduced by the profile-loss formulation.

Detection and mixed-authorship localization. Document-level machine-text detection has been studied through supervised classifiers, likelihood and rank statistics, perturbation tests, rewriting distances, intrinsic-dimension summaries, and watermark-based mechanisms. Representative zero-shot procedures include GLTR-style token-rank diagnostics, DetectGPT and Fast-DetectGPT probability-curvature statistics, Binoculars, and intrinsic-dimension or divergent n-gram tests (Gehrmann et al., 2019; Mitchell et al., 2023; Bao et al., 2024; Hans et al., 2024; Tulchinskii et al., 2023). Supervised approaches train classifiers on paired human and machine corpora or adapt detection rules to a target generator. These methods differ in their assumptions and transfer behavior, but most produce a scalar score or probability that can be used as an input to the present framework. RWCP is detector-agnostic: it changes how ordered scores are segmented, not how a single text unit is scored.

Mixed human–AI writing has motivated sentence-level prediction, token-to-sentence aggregation, window voting, supervised sequence segmentation, and partial-LLM detectors. The CoAuthor data record detailed human interactions with model suggestions and provide a particularly important real-world test bed (Lee et al., 2022). Existing localization procedures often classify each unit independently, aggregate token predictions, or smooth local predictions over a fixed window (Zeng et al., 2024; Zhang et al., 2024; Lei et al., 2025). Change point formulations impose a stronger structural prior: consecutive units share a source until a boundary occurs. This structure reduces isolated label flips and aligns the optimization objective with boundary localization.

The foundational WCP formulation makes a further advance by incorporating heterogeneous score variability. Its upper and lower bounds show that inverse-variance information, rather than raw sentence count, determines localization difficulty. Our work preserves this geometry but replaces the mean-based split statistic by a robust profile-loss gain and extends the stochastic analysis to locally dependent bounded-influence scores.

Change points and robust profile methods. CUSUM statistics are central to univariate mean-change detection (Wang et al., 2020; Yu, 2020). Binary segmentation, wild binary segmentation, and narrowest-over-threshold (NOT) algorithms use multiscale or random intervals to isolate individual changes (Fryzlewicz, 2014; Baranowski et al., 2019). NOT selects the shortest interval whose maximal contrast exceeds a threshold and then recurses, thereby reducing cancellation among multiple changes. Non-asymptotic localization theory typically combines three ingredients: a deterministic contrast margin, a uniform noise bound, and an interval-isolation event.

Our proof follows that architecture in reliability coordinates. The deterministic object is a profile gain rather than a linear contrast; the noise control is based on bounded Huber scores under an explicit dependence-aware Bernstein condition; and random intervals are sampled in cumulative reliability mass. The recursive proof also has to control boundary leakage because an estimated split need not coincide exactly with the true change.

Huber loss interpolates between quadratic and absolute loss and yields bounded score influence (Huber, 1964; Huber & Ronchetti, 2009). Robust mean and location estimators have been developed under finite moments, contamination, and dependence. Median-of-means and Catoni-type estimators offer alternative guarantees, while Huber estimators are especially convenient here because their optimized objective defines a profile likelihood-ratio analogue. Robust change point methods include median and rank contrasts, robustified CUSUMs, M-estimation losses, and procedures designed for heavy-tailed or contaminated sequences (Li & Yu, 2021; Cho & Kirch, 2022).

Profile costs of the form $\min_\theta \sum_i \rho(Y_i - \theta)$, including Huber losses, are established in robust penalized segmentation (**?**). Bounded influence alone does not provide immunity to arbitrary outliers; in particular, an unbounded Huber loss and a bounded robust loss have different contamination behavior. Our guarantees require finite risk, central mass, bounded weights, local jumps, and the stated concentration condition. The specific distinction analyzed here is between a robust center-difference contrast and a reliability-weighted profile-loss gain. The former requires understanding a nonlinear mixed-distribution center at every candidate split. The latter compares optimized risks and converts an incorrect split into a merge-cost problem. This distinction is not cosmetic; it is the main reason a transparent deterministic margin can be proved.

Dependence and calibration. Text scores may be dependent because adjacent sentences share topic, vocabulary, discourse structure, and detector context. Change point theory under mixing and other weak-dependence conditions replaces independent concentration by blocking, coupling, martingale,

or spectral arguments (Doukhan, 1994; Merlevède et al., 2009). Block bootstrap and dependent multiplier methods are widely used to approximate distributions of statistics under serial dependence (Künsch, 1989). In our main theory, dependence enters through an explicit Bernstein condition. Independent and fixed-order dependent sequences have the usual logarithmic complexity, while the geometric beta-mixing specialization retains the published logarithmic penalty. The bootstrap is kept secondary: it is a practical calibration mechanism, while exact recovery is established using an analytic threshold.

B Experimental Protocol and Additional Results

This appendix documents the cached-score inventory, fixed RWCP-R configuration, selection history, and evaluation conventions before presenting the full benchmark comparisons. It then reports recovery and component diagnostics, post-freeze stability, implementation details, and paired uncertainty estimates. All 25 primary settings and 2,690 instances use the same reporting conventions as Section 4.

B.1 Data, comparators, and evaluation scope

The controlled domains are denoted WikiQA, News, and Story following the reference segmentation pipeline (Li et al., 2026); their source corpora are SQuAD, XSum, and WritingPrompts, respectively (Rajpurkar et al., 2016; Narayan et al., 2018; Fan et al., 2018). In the primary suite, single-change and attack data use cached Claude Haiku 4.5 texts (`claude-haiku-4-5`); multiple-change Story also uses `gpt-5-mini`. CoAuthor retains its existing 868/289/290 training/validation/test session partition and three source classes: human, collaborative, and AI (Lee et al., 2022). The evaluated split has 290 sessions and 7,042 sentences.

Table 5: Completed benchmark inventory. Counts are evaluated document instances; transformed copies can share a source. FT is the repository's fine-tuned score and NFT its normalized score. The retained benchmarks use the unknown-count protocol for segmentation.

Experiment	Settings	Instances	Score	Count protocol
Single change	3	300	FT	Unknown K
CoAuthor	1	290	NFT	Unknown K
Multiple changes	10	1,000	NFT	Unknown K
Text attacks	6	600	NFT	Unknown K
LLM proportion	5	500	NFT	Unknown K

Every controlled setting contains 100 records, including WikiQA, News, and Story. In the three Claude WikiQA settings (single change and two attacks), records 80 and 86 each contain one sentence and no reference boundary. We retain these records: the only admissible boundary set is empty, both window errors and count MAE are zero, and empty–empty boundary F1 is one. This is a boundary-only convention, not a claim of correct source classification. It is applied equally to every method. The other 98 records per setting retain their saved predictions. Thus the primary suite contains 2,690 instances; six trivial instances enter reporting only and do not alter historical parameter selection. The multiple-change settings use $K \in \{1, 2, 3, 5, 8\}$ for each generator; the proportion settings use target fractions of 5%, 10%, 20%, 40%, and 80%. We reuse the saved hybrid texts, predictions, and detector scores. Generation revisions, dates, and decoding metadata are incomplete; reproducibility is from these caches, rather than exact regeneration from a model name.

Comparators are TextTiling (Hearst, 1997), direct LLM prediction (LLMPred), independent sentence prediction (SenPred), local majority voting (Voting), VCP, and WCP. VCP/WCP use the reference change-point implementation (Li et al., 2026); sentence prediction and voting follow the localization pipeline (Zhang et al., 2024). PaLD (Lei et al., 2025) has a complete local output only on WikiQA in the single-change benchmark. SegFormer outputs on CoAuthor and VCP outputs for the attack/proportion experiments are unavailable. We retain saved baseline predictions and recompute metrics consistently; RWCP-R is the rerun method. The new core RWCP comparison is also recomputed from the same caches, but uses its own fixed analytic threshold rather than the historical practical RWCP-analytic/block configurations.

For every nontrivial input in the primary suite, core RWCP and RWCP-R consume exactly the WCP cached score sequence and reliability proxies. FT and NFT are distinct score pipelines, and other baselines need not share their representations. Historical WCP re-scores concatenated predicted segments with the language model, whereas RWCP-R labels its states from fitted score centers. Thus final-label metrics compare complete pipelines. Standard WD on raw boundaries and the component ablations provide complementary evidence about segmentation.

B.2 Fixed analytic-threshold core RWCP evaluation

To evaluate the method analyzed in Theorem 3.11, we run the weighted Huber profile gain, reliability-coordinate interval search, analytic threshold, and guarded recursion without the scale estimation, additional multiscale intervals, global penalized fitting, or state sharing of the practical variants. The fixed numerical configuration is $c = 1.345$, capped weights in $[0.2, 5]$, minimum side and guard masses both 1, 500 sampled intervals, $\delta = 0.05$, threshold multiplier $C_r = 1$, dependence multiplier 1, and seed 42. Cached WCP reliability proxies are clipped directly without per-document renormalization. The implementation additionally includes the full document as a deterministic candidate interval and minimizes each Huber loss over the observed score span; these details are outside the theorem's literal fixed-Θ, sampled-interval specification. The run does not verify the unknown signal, spacing, curvature, or dependence constants required by the theorem, and should not be read as a test of its probability guarantee.

Table 6: Fixed analytic-threshold core RWCP versus saved WCP boundaries on the same cached scores. Family means average documents within settings and then settings equally. F1 matches boundaries within one sentence. Zero CP is the fraction with no detected boundary. Loc. coverage is the fraction of documents with at least one true change whose estimated change count is exactly correct. Weighted loc. MAE averages ordered boundary distances d_w only over those eligible documents; – means none is eligible. The conditional localization mean must be interpreted together with coverage. These are descriptive results without newly estimated paired intervals.

	WCP			Core RWCP					
Family	WD	MAE	F1	WD	MAE	F1	Zero CP	Loc. cov.	Weighted loc. MAE
Single change	0.3961	1.637	0.435	0.2315	0.393	0.487	31.3%	61.4%	1.636
CoAuthor	0.4878	7.917	0.302	0.4897	10.159	0.057	82.4%	0.0%	–
Multiple changes	0.4363	2.130	0.205	0.4108	3.716	0.026	93.1%	1.0%	0.892
Text attacks	0.3609	0.998	0.272	0.3369	0.940	0.038	94.7%	5.4%	2.788
LLM proportion	0.3538	1.360	0.220	0.2085	0.936	0.037	92.8%	6.4%	4.395

Across the 25 settings, the core implementation has lower family-macro WD than WCP, but its count MAE and F1 are worse overall. On single-change texts, the 61.4% localization coverage is higher than WCP's 30.9%, yet the conditional weighted localization error is 1.636 for core RWCP versus 0.821 for WCP; these means use different eligible subsets and cannot establish a per-boundary improvement. Only 1.0% of multiple-change documents have the correct positive change count under core RWCP. The apparent conditional weighted error of 0.892 in that family rests on just 10 eligible documents. The zero-boundary rates and low coverage limit any empirical claim of exact recovery on these cached detector scores.

Table 7: Per-setting standard WD for WCP and fixed analytic-threshold core RWCP, with core count MAE, boundary F1 at one-sentence tolerance, and no-boundary fraction. Each setting contains 100 documents except CoAuthor (290). The six one-sentence WikiQA records follow the common empty-boundary convention.

Family	Setting	WCP WD	Core WD	Core MAE	Core F1	Zero CP
Single	WikiQA	0.284	0.279	0.580	0.380	60%
Single	News	0.370	0.119	0.130	0.760	8%
Single	Story	0.534	0.296	0.470	0.320	26%
CoAuthor	Sessions	0.488	0.490	10.159	0.057	82%

Family	Setting	WCP WD	Core WD	Core MAE	Core F1	Zero CP
Multiple	Claude $K = 1$	0.400	0.319	0.920	0.070	92%
Multiple	Claude $K = 2$	0.404	0.393	1.900	0.047	91%
Multiple	Claude $K = 3$	0.425	0.419	2.840	0.051	88%
Multiple	Claude $K = 5$	0.425	0.447	4.830	0.035	87%
Multiple	Claude $K = 8$	0.430	0.460	7.820	0.022	86%
Multiple	GPT-5-mini $K = 1$	0.434	0.333	1.000	0.000	100%
Multiple	GPT-5-mini $K = 2$	0.430	0.395	1.970	0.020	97%
Multiple	GPT-5-mini $K = 3$	0.470	0.424	2.970	0.009	98%
Multiple	GPT-5-mini $K = 5$	0.475	0.453	4.960	0.006	97%
Multiple	GPT-5-mini $K = 8$	0.471	0.464	7.950	0.002	95%
Attack	WikiQA decoh.	0.340	0.365	0.980	0.020	100%
Attack	News decoh.	0.407	0.336	1.000	0.000	100%
Attack	Story decoh.	0.398	0.321	0.860	0.060	86%
Attack	WikiQA para.	0.266	0.361	0.970	0.030	99%
Attack	News para.	0.326	0.314	0.930	0.070	93%
Attack	Story para.	0.428	0.325	0.900	0.050	90%
Proportion	5%	0.321	0.076	0.980	0.007	97%
Proportion	10%	0.338	0.137	0.970	0.017	96%
Proportion	20%	0.375	0.259	0.920	0.040	91%
Proportion	40%	0.404	0.301	0.870	0.090	86%
Proportion	80%	0.330	0.269	0.940	0.030	94%

B.3 Practical RWCP configuration

RWCP-R implements Section 2.2, with three declared source states on CoAuthor and two elsewhere. Given positive cached reliability proxies r_i, we compute

$$\widetilde{w}_i = \mathrm{clip}_{[0.05,10]}\left(\frac{r_i^{\eta}}{N^{-1}\sum_j r_j^{\eta}}\right), \qquad \eta = 0.5,$$

and do not renormalize after clipping. The cache proxy is n_i^2/v_i: n_i is the scoring tokenizer's unit token count and v_i the NFT score-variance proxy; FT uses $v_i = 1$. Compression tempers extreme reliability ratios without discarding their ordering. All settings use the same exponent.

Scores are centered by their median and divided by an initial robust scale. For adjacent differences, let

$$D_i = \frac{Y_{i+1} - Y_i}{\sqrt{1/\widetilde{w}_i + 1/\widetilde{w}_{i+1}}}, \qquad s_D = \frac{\mathrm{median}_i\,|D_i - \mathrm{median}(D)|}{0.67448975}.$$

The initial scale is the maximum of s_D, 0.05 times the Gaussian-scaled MAD of score levels, and 10^{-6} times the score range; a constant sequence uses scale one. The fitted dimensionless scale in (6) is constrained by $\sigma \geq 0.25$.

The frozen main configuration uses Huber cutoff $c = 1.5$ and transition penalty $\lambda = 4$. Initial centers are equally spaced quantiles between $(q, 1-q)$, for $q \in \{0.05, 0.20, 0.35\}$, combined with initial scales one and 0.25, giving six deterministic starts. Each start runs at most 50 alternating updates, stopping when the objective change is below $10^{-8}(1 + |\mathcal{J}|)$. Center and scale bisection use tolerance 10^{-8}. We retain the smallest attained objective, breaking exact ties by fewer transitions. Empty states retain their previous centers. Fitted centers are ranked in ascending order to assign source-state labels; this assumes score ordering agrees with source semantics. The smallest allowed state run is one observation; no reference boundary, target LLM proportion, generator identity, or reference change count enters an unknown-count prediction.

The implementation uses at most 64 bisection iterations per conditional fit with absolute tolerance 10^{-8}. For B bisection steps and T alternating iterations, six starts cost $O(6TN(S^2 + SB))$ time and $O(NS)$ decoding memory. The scale ablation fixes $\sigma = 1$ and consequently uses only the three center initializations. The dynamic program is exact conditional on centers and scale; the overall fit is a deterministic multistart approximation.

Algorithm 1 Recurrent profile decoding (RWCP-R)

Require: Scores, reliability proxies, declared state count S, and frozen $c, \eta, \lambda, \sigma_{\min}$.
If $N = 1$, return no boundary (the boundary-only evaluation convention).
Standardize scores and construct capped weights as specified above.
for each of the six quantile/scale initializations **do**
    for $t = 1, \ldots, 50$ **do**
        Decode z by Viterbi with Huber emissions and penalty λ.
        Fit each occupied center by weighted Huber bisection; retain empty centers.
        Let $F(\sigma) = \sum_i \psi_c(r_i(\sigma)) r_i(\sigma)$.
        Set $\sigma = \sigma_{\min}$ if $F(\sigma_{\min}) \le N$; otherwise solve $F(\sigma) = N$.
        Stop if the objective change is below $10^{-8}(1 + |\mathcal{J}|)$.
    end for
end for
Keep the smallest attained objective; break exact ties by fewer transitions.
Rank centers to label states and return $\{i : z_i \neq z_{i+1}\}$.

For continuity, RWCP-analytic and RWCP-block denote the preceding `analytic_block_v3` implementation, using $c = 3$, clipped mean-one weights in $[0.02, 10]$, three refinement passes, and common threshold multiplier 0.85. These practical core variants also use implementation refinements beyond the unmodified analytic-threshold theorem. Their existing results are retained in Tables 15 and 17. The five retained comparisons use RWCP-R throughout. The historical core configurations fix 500 sampled intervals (with exhaustive intervals for $N \le 24$), minimum side mass and guard mass one, nominal level 0.05, dependence factor one, seed 42, and optimization tolerance 10^{-6}. Block calibration uses length four and 499 multiplier draws; its threshold mixes the bootstrap value with the analytic reference at weight 0.35 on the latter and is capped at 1.5 times that reference. Both variants use curvature floor 0.05, curvature shrinkage 0.5, and the common threshold multiplier 0.85. These empirical values are not asserted to satisfy the theorem's unknown signal/spacing requirements.

B.4 Development path and leakage controls

Configuration selection minimized the family-balanced development mean of

$$\frac{1}{2}\mathrm{WD}_d + \frac{1}{2}\min\{\mathrm{WD}_{\sup}(d), 1\} + \frac{1}{4}\frac{|K_d - \widehat{K}_d|}{K_d + 1} + \frac{1}{4}\{1 - \mathrm{F1}_{d,h=1}\}.$$

The search was shared across all reported settings rather than repeated separately for each dataset. Table 8 records the adaptive sequence. The final recurrent sweep was the Cartesian product $c \in \{1.5, 3\}$, $\eta \in \{0.25, 0.5, 1\}$, $\lambda \in \{0.25, 0.5, 1, 2, 4\}$, and $\sigma_{\min} \in \{0.25, 0.5\}$, giving 60 candidates. The selected point is the frozen configuration used in the primary results and the reference point for the component ablations.

Table 8: Adaptive development path under the original broader protocol. The objective is the criterion above averaged first within families and then across families; lower is better. These are development-selection values, not test-set performance estimates.

Stage	Candidates	Main change	Best dev. objective
Existing RWCP-block	fixed	NOT with block calibration	0.62712
Global profile study	45	global Huber profile and state pruning	0.61382
Recurrent RWCP-R	60	recurrent centers and fitted scale	**0.58506**

The historical nontrivial subset contains 643 linked source components, with 716 instances assigned to development and 1,968 to the internal check by a fixed hash. Six singleton instances restored for complete 100-record reporting do not enter either split or parameter selection. RWCP-R parameters and all component ablations were frozen before that check. No reference boundary, reference change count, target LLM proportion, or generator identity enters an unknown-count prediction, and paired intervals resample source components rather than individual transformed records. The search was adaptive rather than preregistered, and the internal check still uses benchmark sources that influenced earlier development; it therefore supports a post-freeze stability check, not an independent confirmatory claim.

B.5 METRICS, SELECTION, AND PROVENANCE

Standard boundary WindowDiff (Pevzner & Hearst, 2002), defined in (21), compares raw detected boundary counts in windows of width

$$k_d = \max\left\{1, \operatorname{round}\left(\frac{N_d}{2(K_d+1)}\right)\right\}, \tag{18}$$

with ties-to-even rounding. It lies in $[0, 1]$. For comparison with the inherited tables, $\mathrm{WD}_{\mathrm{sup}}$ instead averages the absolute discrepancy between reference and final-label boundary counts, using width $\max\{1, \lfloor N_d/[2(K_d+1)] \rfloor\}$. It can exceed one and is not standard WD. Old aggregate fields are not mixed with new RWCP values.

Raw count error is $\mathrm{CE}_d = K_d - \widehat{K}_d$, positive for missed changes, and count MAE averages $|K_d - \widehat{K}_d|$ over documents. CoAuthor also retains its historical final-label error $\mathrm{CE}_{\mathrm{label}}$. Boundary F1 uses one-to-one matching of estimated and reference boundaries within tolerance $h = 1$ observation; $h = 0$ is an exact-match diagnostic. Two empty sets receive F1 one and exactly one empty set receives zero. For singleton documents, which have no evaluable window, both window errors are defined as zero. The zero-detection fraction includes the retained singletons; outside these known no-change records, a high value diagnoses missed boundaries. Table boldface is computed before rounding and does not denote significance.

The reported suite focuses on sentence-level segmentation with unknown change counts. This scope was narrowed after inspecting the broader development results, so its comparisons are retrospective. The selected configuration and predictions were retained without retuning for this subset.

The development objective, adaptive search path, source-group split, freezing decisions, and leakage controls are summarized in Appendix B.4; this appendix retains the complete metric definitions and benchmark-specific results. Source IDs are incomplete, so component linkage reduces obvious reuse but does not establish independence.

Pointwise paired 95% intervals against WCP use 2,000 bootstrap repetitions, resampling source components within each setting. We report them in Appendix B.8.3, without multiplicity correction. Family summaries average settings equally; the component summary then averages the five retained families equally.

B.6 FIVE BENCHMARK COMPARISONS

Single-change localization. In the primary Claude evaluation, RWCP-R has the lowest $\mathrm{WD}_{\mathrm{sup}}$ on all three domains (Table 9): 0.184 on WikiQA, 0.097 on News, and 0.235 on Story. Their strongest available baseline values are 0.200, 0.123, and 0.306. Standard WD averaged over the domains falls from 0.2441 for RWCP-block to 0.1895 for RWCP-R, and count MAE falls from 0.490 to 0.340. Paired standard-WD intervals against WCP exclude zero on all three domains, showing a consistent pipeline-level gain in the controlled Claude single-change setting. This comparison alone does not isolate Huber loss from state sharing, scale fitting, or penalty selection.

The GPT-5-mini aggregate extension is shown separately in Table 10. RWCP-R has lower legacy window error on News and Story, while VCP is lower on WikiQA. These supplied aggregates do not enter family means, ablations, win counts, or paired intervals for the primary study.

Table 9: Primary Claude Haiku 4.5 single-change results, 100 records each in WikiQA, News, and Story. Values are document means ± sample standard deviations, recomputed with the singleton boundary convention. Bold marks the lowest unrounded $\mathrm{WD}_{\mathrm{sup}}$ or smallest absolute mean CE, not significance. PaLD has outputs only for WikiQA.

Method	WikiQA		News		Story	
	$\mathrm{WD}_{\mathrm{sup}} \downarrow$	$\mathrm{CE} \to 0$	$\mathrm{WD}_{\mathrm{sup}} \downarrow$	$\mathrm{CE} \to 0$	$\mathrm{WD}_{\mathrm{sup}} \downarrow$	$\mathrm{CE} \to 0$
TextTiling	0.388±0.267	-0.09±0.84	1.066±0.244	-4.18±0.99	0.850±0.233	-3.19±0.94
LLMPred	0.362±0.204	0.41±0.96	0.507±0.456	-0.81±1.94	0.740±0.713	-2.02±3.26
PaLD	0.446±0.330	-1.28±1.23	–	–	–	–
SenPred	0.271±0.291	-0.90±1.11	1.663±0.829	-6.74±3.18	2.972±1.071	-12.51±4.29
Voting	0.282±0.277	**0.00±0.40**	0.131±0.213	-0.32±0.80	0.635±0.549	-2.39±2.24
VCP	0.200±0.187	0.20±0.68	0.124±0.186	-0.99±1.43	0.306±0.269	-1.02±1.63
WCP	0.200±0.190	0.07±0.81	0.123±0.219	-1.73±1.87	0.372±0.307	-2.55±2.11
RWCP-R	**0.184±0.170**	0.47±0.50	**0.097±0.142**	**0.03±0.30**	**0.235±0.163**	**0.34±0.59**

Table 10: Supplemental GPT-5-mini single-change panel: 100 records per domain, as specified by the dataset inventory. Entries preserve the supplied aggregate means ± sample standard deviations. This panel is outside the primary 25-setting evaluation and its paired comparisons; document-level outputs and matching selection provenance are not included in that evaluation bundle.

Method	WikiQA		News		Story	
	$\mathrm{WD}_{\mathrm{sup}} \downarrow$	$\mathrm{CE} \to 0$	$\mathrm{WD}_{\mathrm{sup}} \downarrow$	$\mathrm{CE} \to 0$	$\mathrm{WD}_{\mathrm{sup}} \downarrow$	$\mathrm{CE} \to 0$
TextTiling	0.36±0.24	**-0.02±0.81**	1.32±0.30	-5.19±0.97	1.11±0.30	-4.39±1.06
LLMPred	0.44±0.20	0.09±0.64	0.50±0.33	**-0.59±1.46**	0.52±0.46	-0.84±2.16
PaLD	0.50±0.38	-1.41±1.46	1.76±0.75	-6.75±3.37	2.19±0.89	-8.47±3.82
SenPred	0.52±0.38	-1.62±1.51	3.32±0.85	-13.36±3.09	3.61±1.05	-14.49±3.89
Voting	0.39±0.28	-0.10±0.77	0.89±0.50	-3.17±2.15	1.12±0.46	-4.14±2.00
VCP	**0.25±0.18**	0.39±0.74	0.41±0.26	-0.83±1.77	0.52±0.27	-0.69±1.88
WCP	0.27±0.19	0.26±0.76	0.44±0.26	-1.16±1.87	0.57±0.29	-1.85±1.99
RWCP-R	0.26±0.17	0.64±0.54	**0.32±0.12**	0.62±0.55	**0.35±0.12**	**0.64±0.59**

Real co-authoring. On CoAuthor (Table 11), RWCP-R has the smallest $\mathrm{WD}_{\mathrm{sup}}$ mean, 0.4780 versus 0.4804 for VCP. Standard WD is 0.4869 versus 0.4878 for WCP. The table reports count and boundary diagnostics alongside the window-error gains, allowing the recurrent objective to be compared with the historical labeling pipeline.

Table 11: CoAuthor: 290 sessions and 7,042 sentences. $\mathrm{CE}_{\mathrm{label}}$ is the historical final-label count error; raw CE, count MAE, and Zero CP use detected boundaries. RWCP-R has the smallest $\mathrm{WD}_{\mathrm{sup}}$ mean. Zero CP is descriptive and is not ranked. SegFormer has no local output.

Method	Final labels		Raw boundaries		
	$\mathrm{WD}_{\mathrm{sup}} \downarrow$	$\mathrm{CE}_{\mathrm{label}} \to 0$	$\mathrm{CE} \to 0$	MAE↓	Zero CP
TextTiling	0.551	7.31	7.31	7.62	0.0%
LLMPred	0.516	8.30	8.30	8.51	19.0%
SenPred	0.809	-4.16	-4.16	5.58	0.0%
Voting	0.647	**3.29**	**3.29**	**5.27**	0.0%
VCP	0.480	9.42	9.37	9.40	61.7%
WCP	0.483	7.98	7.72	7.92	15.9%
RWCP-R	**0.478**	10.17	10.17	10.19	84.1%

Multiple changes. RWCP-R leads in five of ten $\mathrm{WD}_{\mathrm{sup}}$ settings (Table 12): Claude $K = 1$ and GPT-5-mini $K = 1, 3, 5, 8$. VCP leads at Claude $K = 2, 3, 5$, Voting at Claude $K = 8$, and WCP at GPT-5-mini $K = 2$. Against the previous block configuration, mean standard WD improves from

0.4112 to 0.3993 and count MAE from 3.628 to 3.233. The full recurrent pipeline improves these averages over the historical block configuration, but this comparison does not isolate the effect of shared states. WCP still has lower count MAE and higher F1 in this family.

Table 12: Multiple-change Story results: 100 documents per generator and true K. The displayed count is not provided to unknown-count detectors. Entries are mean $\mathrm{WD}_{\mathrm{sup}}$ and raw CE. RWCP-R leads in five of ten WD settings.

Method	$K=1$		$K=2$		$K=3$		$K=5$		$K=8$	
	$\mathrm{WD}_{\mathrm{sup}}\downarrow$	$\mathrm{CE}\to 0$	$\mathrm{WD}_{\mathrm{sup}}\downarrow$	$\mathrm{CE}\to 0$	$\mathrm{WD}_{\mathrm{sup}}\downarrow$	$\mathrm{CE}\to 0$	$\mathrm{WD}_{\mathrm{sup}}\downarrow$	$\mathrm{CE}\to 0$	$\mathrm{WD}_{\mathrm{sup}}\downarrow$	$\mathrm{CE}\to 0$
Claude Haiku 4.5										
TextTiling	0.898	-3.35	0.702	-3.57	0.711	-4.49	0.588	-4.66	0.483	-2.90
LLMPred	0.714	-1.99	0.689	-1.71	0.733	-1.83	0.624	0.42	0.504	4.84
SenPred	2.711	-11.28	2.712	-16.79	2.619	-22.17	2.183	-27.81	1.583	-29.68
Voting	0.476	-1.85	0.464	-2.15	0.475	-3.22	0.459	-4.29	**0.348**	-3.30
VCP	0.293	-1.13	**0.319**	-1.02	**0.315**	-1.32	**0.351**	-1.53	0.376	**0.35**
WCP	0.313	**-0.52**	0.337	**-0.19**	0.339	**-0.11**	0.369	**0.39**	0.383	2.85
RWCP-R	**0.273**	0.58	0.360	1.44	0.354	2.10	0.397	4.26	0.409	7.21
GPT-5-mini										
TextTiling	1.120	-4.39	0.905	-4.66	0.882	-6.12	0.734	-7.01	0.536	-4.70
LLMPred	0.645	-1.52	0.778	-2.29	0.696	-1.16	0.681	**-0.16**	0.542	3.99
SenPred	3.444	-13.56	3.275	-19.45	3.225	-25.81	2.685	-33.02	1.894	-35.41
Voting	1.128	-4.09	0.997	-5.25	1.018	-6.78	0.874	-8.29	0.675	-7.04
VCP	0.445	-0.95	0.458	-1.05	0.451	-0.82	0.462	-0.35	0.454	**2.03**
WCP	0.423	**-0.38**	**0.404**	**0.01**	0.450	**0.31**	0.438	1.26	0.438	4.10
RWCP-R	**0.342**	0.79	0.419	1.51	**0.419**	2.42	**0.427**	4.28	**0.429**	7.22

Text perturbations. RWCP-R has the lowest $\mathrm{WD}_{\mathrm{sup}}$ in five of six attack settings (Table 13); Voting remains best on paraphrased News. Standard WD falls from 0.3273 for RWCP-block to 0.2676, and count MAE from 0.785 to 0.602. Unlike the earlier block configuration, RWCP-R improves standard WD against WCP in both attacked WikiQA conditions; all six pointwise paired intervals against WCP are negative across the cached perturbations.

Table 13: Text-attack results: 100 records in each of WikiQA, News, and Story per attack. Predictions for nontrivial inputs are unchanged; singleton boundary conventions match Table 9. RWCP-R has the lowest $\mathrm{WD}_{\mathrm{sup}}$ in five of six settings; Voting leads on paraphrased News.

Method	WikiQA		News		Story	
	$\mathrm{WD}_{\mathrm{sup}}\downarrow$	$\mathrm{CE}\to 0$	$\mathrm{WD}_{\mathrm{sup}}\downarrow$	$\mathrm{CE}\to 0$	$\mathrm{WD}_{\mathrm{sup}}\downarrow$	$\mathrm{CE}\to 0$
Decoherence						
TextTiling	0.388	-0.09	1.066	-4.18	0.844	-3.17
LLMPred	0.352	0.37	0.604	-1.22	0.957	-2.96
SenPred	0.383	-1.30	2.573	-10.58	3.119	-12.89
Voting	0.337	**-0.03**	0.491	-1.75	0.993	-3.82
WCP	0.254	0.44	0.328	**-0.59**	0.364	**-0.45**
RWCP-R	**0.226**	0.65	**0.324**	0.60	**0.328**	0.59
Paraphrasing						
TextTiling	0.389	-0.06	1.049	-4.16	0.839	-3.21
LLMPred	0.368	0.30	0.445	-0.44	0.737	-2.13
SenPred	0.294	-0.96	1.590	-6.53	2.857	-12.08
Voting	0.264	**0.04**	**0.107**	-0.24	0.581	-2.23
WCP	0.187	0.25	0.167	-1.05	0.336	-0.75
RWCP-R	**0.148**	0.37	0.141	**0.12**	**0.303**	**0.52**

LLM-written proportion. RWCP-R leads in $\mathrm{WD}_{\mathrm{sup}}$ at 5%, 10%, 40%, and 80%, with WCP leading at 20% (Table 14). The setting-average standard WD is 0.2119, slightly worse than RWCP-

block's 0.2084, while count MAE improves from 0.874 for RWCP-block to 0.736. The consistent gains across four target proportions indicate that the recurrent profile remains effective as the LLM-written fraction varies.

Table 14: Story with varying target LLM-written fractions, 100 documents per setting. RWCP-R leads in $\mathrm{WD}_{\mathrm{sup}}$ at 5%, 10%, 40%, and 80%; WCP leads at 20%. Neither the target fraction nor the true boundary is supplied to RWCP-R.

Method	5%		10%		20%		40%		80%	
	$\mathrm{WD}_{\mathrm{sup}} \downarrow$	$\mathrm{CE} \to 0$	$\mathrm{WD}_{\mathrm{sup}} \downarrow$	$\mathrm{CE} \to 0$	$\mathrm{WD}_{\mathrm{sup}} \downarrow$	$\mathrm{CE} \to 0$	$\mathrm{WD}_{\mathrm{sup}} \downarrow$	$\mathrm{CE} \to 0$	$\mathrm{WD}_{\mathrm{sup}} \downarrow$	$\mathrm{CE} \to 0$
TextTiling	1.166	-3.92	1.155	-4.15	1.048	-4.15	1.005	-3.68	0.766	-2.75
LLMPred	0.620	-1.50	0.700	-1.70	0.698	-1.26	0.692	-1.26	0.525	-1.06
SenPred	4.500	-16.48	4.264	-15.64	3.880	-14.77	3.233	-13.18	3.134	-12.43
Voting	1.503	-5.04	0.768	-2.64	0.671	-2.40	0.629	-2.46	0.943	-3.28
WCP	0.213	-1.02	0.235	-0.95	**0.228**	-1.13	0.289	-0.82	0.275	**-0.10**
RWCP-R	**0.111**	**0.79**	**0.185**	**0.49**	0.243	**0.48**	**0.277**	**0.52**	**0.271**	0.72

B.7 Recovery diagnostics and component ablations

B.7.1 Family-wise comparison of RWCP variants

Table 15 reports the family-wise results for the preceding analytic and block-calibrated variants alongside RWCP-R. RWCP-R has the lowest WD in four families, while the block variant is lower in the LLM-proportion family. The comparisons retain historical configurations and do not isolate state sharing.

Table 15: Standard raw-boundary WD, equally averaged over settings within families. Best baseline requires complete family coverage. Analytic and Block are historical fixed configurations; their differences from RWCP-R include search, scale, and other settings. This comparison is not an isolated ablation of recurrence.

Family	Best baseline		Analytic	Block	RWCP-R
	Method	WD	WD	WD	WD
Single change	Voting	0.2725	0.2853	0.2441	**0.1895**
CoAuthor	WCP	0.4878	0.4922	0.4945	**0.4869**
Multiple changes	WCP	0.4363	0.4094	0.4112	**0.3993**
Text attacks	Voting	0.3514	0.3307	0.3273	**0.2676**
LLM proportion	WCP	0.3538	0.2155	**0.2084**	0.2119

B.7.2 Recovery stress tests

The main text emphasizes WD and signed CE. Table 16 additionally reports absolute count error, matched-boundary F1, the no-boundary fraction, and the WD of a predictor that returns no boundary. These diagnostics distinguish a lower window error from successful boundary recovery and expose count errors that signed averaging can cancel.

Single changes and text attacks improve on the no-boundary predictor by 0.1552 and 0.0772 WD and improve boundary F1 relative to WCP by 0.110 and 0.028. CoAuthor and LLM proportion are within 0.0024 WD of the no-boundary predictor and return no boundary in 84.1% and 66.8% of instances. Their F1 differences against WCP are -0.244 and -0.069; multiple-change data improve modestly on the null WD but lose 0.108 F1. Across the five families, count MAE is 3.019 for RWCP-R versus 2.808 for WCP, and F1 is 0.230 versus 0.287. Thus the primary window-error improvement is accompanied by worse aggregate count and boundary diagnostics. Table 19 gives the paired intervals.

B.7.3 Complete component diagnostics

The full component comparison in Table 17 supplements the WD/CE presentation in Table 4. Scale adaptation improves both window errors, count MAE, and boundary F1 at these configurations.

Table 16: Raw-boundary recovery diagnostics, equally averaged over settings. F1 matches boundaries within one observation; zero CP is the no-boundary fraction and Null WD is the no-boundary predictor. The complete-record convention applies throughout.

Family	WCP		RWCP-R			No boundary
	MAE	F1	MAE	F1	Zero CP	WD
Single change	1.637	0.435	0.340	0.544	31.7%	0.3448
CoAuthor	7.917	0.302	10.186	0.058	84.1%	0.4893
Multiple changes	2.130	0.205	3.233	0.097	63.9%	0.4181
Text attacks	0.998	0.272	0.602	0.300	54.5%	0.3448
LLM proportion	1.360	0.220	0.736	0.151	66.8%	0.2141

Uniform weights improve count MAE and both F1 variants while nearly matching standard WD; the quadratic approximation also improves count MAE and F1 relative to the full method. These fixed-configuration comparisons support metric-specific conclusions rather than a claim that every component improves localization.

Table 17: Frozen component ablations, averaged equally over settings and then five families. Fixed scale uses $\sigma = 1$; uniform weights uses exponent zero; the quadratic approximation uses $c = 10^6$. No variant is retuned. Bold marks the best unrounded value per column; it does not indicate significance.

Variant	Window error		Count error	Boundary F1↑	
	$\mathrm{WD}_{\mathrm{sup}}\downarrow$	WD↓	MAE↓	$h=0$	$h=1$
RWCP-analytic	0.3233	0.3466	3.155	0.115	0.176
RWCP-block	0.3158	0.3371	3.200	0.111	0.169
RWCP-R	**0.2990**	**0.3110**	3.019	0.163	0.230
R: fixed scale	0.3038	0.3170	3.071	0.148	0.211
R: uniform weights	0.3058	0.3118	**2.917**	**0.189**	**0.261**
R: quadratic loss	0.3130	0.3179	2.945	0.166	0.239

B.7.4 Post-freeze internal stability

Table 18 reports the historical source-group development and internal-check split. Against the block variant, family-macro WD changes from 0.3374 to 0.2990 on development and from 0.3380 to 0.3165 on the internal check. RWCP-R improves four of five check families; LLM proportion is the exception. These sources had already influenced development and do not constitute a fresh independent test. Code checks verify Huber interval fits, fixed-center decoding against exhaustive enumeration, nonincreasing alternating objectives, score/weight unit invariance, known-count behavior, and boundary/label consistency. Production reruns reproduce the frozen predictions across all five families. Runs use CPU with optional Numba acceleration and cached scores; timing excludes model inference and data generation. Fresh-source evaluation, no-change error calibration on actual detector scores, and matched whole-segment labeling remain necessary for stronger generalization and mechanism claims.

Table 18: Post-freeze stability by the existing source-group split. Development (716 nontrivial instances) supplies parameter selection; internal check (1,968 nontrivial instances) follows configuration freezing. Six reporting-only singletons are outside these historical splits. Both subsets come from benchmarks previously used in research development, and the latter is not a new independent test set. Entries are standard WD.

Experiment	RWCP-block		RWCP-R	
	Dev. WD	Check WD	Dev. WD	Check WD
Single change	0.2501	0.2445	0.1959	0.1895
CoAuthor	0.4803	0.4997	0.4729	0.4921
Multiple changes	0.4075	0.4125	0.3861	0.4036
Text attacks	0.3352	0.3269	0.2573	0.2749
LLM proportion	0.2138	0.2064	0.1827	0.2224
Family macro	0.3374	0.3380	0.2990	0.3165

B.8 Algorithms and Experimental Details

This appendix describes the core Huber solver and block calibration, followed by the metrics used in the completed cached-score evaluation. Algorithm 1 specifies the distinct recurrent solver used for the main empirical results.

One-dimensional Huber optimization. For interval A, define the monotone score

$$G_A(\theta) = \sum_{i \in A} \sqrt{w_i}\, \psi_c\{\sqrt{w_i}(Y_i - \theta)\}.$$

Because ψ_c is nondecreasing in its argument and $Y_i - \theta$ decreases with θ, G_A is nonincreasing. An interior minimizer satisfies $G_A(\theta) = 0$. If $G_A(\theta) > 0$ throughout Θ, the minimizer is the right endpoint; if it is negative throughout, the minimizer is the left endpoint.

Proposition B.1 (Optimization accuracy)**.** *Suppose the empirical loss on A is αS_A^w-strongly convex in a neighborhood containing the exact minimizer $\widehat{\theta}_A$ and numerical solution $\widetilde{\theta}_A$, and suppose $\widehat{\theta}_A \in \mathrm{int}(\Theta)$. If $|\widetilde{\theta}_A - \widehat{\theta}_A| \le \epsilon_{\mathrm{opt}}$, then*

$$0 \le \widehat{\mathcal{L}}_A(\widetilde{\theta}_A) - \widehat{\mathcal{L}}_A(\widehat{\theta}_A) \le \frac{1}{2} S_A^w \epsilon_{\mathrm{opt}}^2.$$

For a profile gain, the total numerical error is at most

$$2 S_{s:e}^w \epsilon_{\mathrm{opt}}^2. \tag{19}$$

Proof. The Huber empirical loss has gradient Lipschitz constant S_A^w by Lemma C.10. Since the exact minimizer has zero subgradient in the interior, smoothness gives the first upper bound. The profile gain contains one parent and two child optimized losses. Their masses sum to $2S_{s:e}^w$ after applying the factor two in (3), which yields the conservative bound (19). □

If the constrained optimum is a boundary point, Algorithm 2 returns that endpoint exactly from the score-sign check; the quadratic numerical-error proposition is used only on the interior branch.

To keep numerical error below the statistical threshold, it is enough to choose

$$\epsilon_{\mathrm{opt}} \le c_{\mathrm{opt}} \sqrt{\frac{\Lambda_N}{W_N}}.$$

In practice a fixed tolerance such as 10^{-6} is usually much smaller.

Warm starts and cached evaluations. For neighboring candidate splits b and $b+1$, the left and right intervals differ by one observation. Their Huber minimizers typically change smoothly. Use $\widehat{\theta}_{s:b}$ to initialize the solve for $[s, b+1]$ and $\widehat{\theta}_{b+1:e}$ for $[b+2, e]$. A safeguarded Newton step is

$$\theta^{(r+1)} = \Pi_\Theta \left[\theta^{(r)} + \frac{G_A(\theta^{(r)})}{H_A(\theta^{(r)}) \vee h_{\min}} \right],$$

Algorithm 2 Bisection for an interval Huber minimizer

Require: Interval A, scores, weights, c, parameter range $[L, U]$, tolerance ϵ_{opt}.

if $G_A(L) \le 0$ **then**
  return L.
else if $G_A(U) \ge 0$ **then**
  return U.
end if
while $U - L > \epsilon_{\text{opt}}$ **do**
  $M \leftarrow (L + U)/2$.
  if $G_A(M) > 0$ **then**
    $L \leftarrow M$.
  else
    $U \leftarrow M$.
  end if
end while
return $(L + U)/2$.

where H_A is defined in (26). If the proposed step does not reduce the loss, revert to bisection. This hybrid retains deterministic convergence while exploiting local quadratic behavior.

For a fixed iterate θ, the score and Hessian can be evaluated from prefix sums of

$$a_i(\theta) = \sqrt{w_i}\psi_c\{\sqrt{w_i}(Y_i - \theta)\}, \qquad h_i(\theta) = w_i \mathbf{1}\{|\sqrt{w_i}(Y_i - \theta)| < c\}.$$

Because these quantities depend on θ, caching is most effective within batches of nearby intervals sharing warm starts.

Complexity accounting. Let $L_m = e_m - s_m + 1$ be the length of random interval m and let J be the average number of score evaluations required by the one-dimensional solver. A direct scan has cost

$$O\left(J \sum_{m=1}^{M} L_m^2\right)$$

because each of $O(L_m)$ splits requires losses over $O(L_m)$ observations. With cached interval data and warm-started score evaluations, the practical cost is closer to

$$O\left(J_{\text{warm}} \sum_{m=1}^{M} L_m + \sum_{m=1}^{M} L_m \log L_m\right),$$

although this is an implementation-dependent empirical complexity rather than a proved worst-case bound. Memory is $O(N + M)$ excluding detector inference. Block calibration multiplies the segmentation-statistic cost by approximately Q, but repetitions are embarrassingly parallel.

Weighted interval generation. Generate cumulative weights once and use binary search in $W(1), \ldots, W(N)$ to map a uniform coordinate to an index. One interval costs $O(\log N)$ time, so pre-sampling M intervals costs $O(M \log N)$. Deduplicate repeated endpoint pairs before profile evaluation. The random seed and the final endpoint list should be stored to make every comparison deterministic across segmentation methods.

One-step multiplier profile approximation. For a homogeneous pilot interval A, the first-order Huber estimator expansion is

$$\widehat{\theta}_A - \theta_A \approx \frac{\sum_{i \in A} \zeta_i}{H_A(\theta_A)},$$

where ζ_i is the centered score. In a bootstrap repetition, replace the score sum by $\sum_{i \in A} \zeta_i^\star$ and curvature by its pilot estimate. The approximate quadratic profile gain is then

$$\Gamma_{s,e}^{\star,\text{1step}}(b) = \frac{\left(H_{s:b}^{-1} \sum_{i=s}^{b} \zeta_i^\star - H_{b+1:e}^{-1} \sum_{i=b+1}^{e} \zeta_i^\star\right)^2}{H_{s:b}^{-1} + H_{b+1:e}^{-1}}, \tag{20}$$

where $H_{a:b}$ denotes the pilot Hessian sum. Formula (20) is the robust analogue of a squared standardized difference. It is used only for calibration, not in the main estimator.

B.8.1 METRIC DEFINITIONS AND REPRODUCIBILITY

WindowDiff and boundary metrics. Let $C_i^{(r)}$ and $C_i^{(h)}$ be the true and estimated numbers of boundaries in window $[i, i+k]$. WindowDiff is

$$\mathrm{WD} = \frac{1}{N-k} \sum_{i=1}^{N-k} \mathbf{1}\{C_i^{(r)} \neq C_i^{(h)}\}. \tag{21}$$

For document d, use the pre-specified width $k = k_d$ in (18); this keeps the implementation identical across methods while adapting to reference segment length. Count Error is

$$\mathrm{CE} = K - \widehat{K}.$$

Match an estimate to a true boundary when $|\widehat{\tau} - \tau| \leq h$, using maximum bipartite matching to avoid multiple credit. The completed RWCP-R evaluation reports F1 for $h = 1$ observation and the exact-match case $h = 0$. Precision, recall, and F1 follow from matched, unmatched estimated, and unmatched true boundaries.

The empirical evidence consists of saved score streams, predictions, source-group assignments, selected configurations, and metric recomputation. The reporting correction includes every controlled record and explicitly records the singleton boundary convention. It preserves the historical development/check memberships and all nontrivial predictions. Generation revisions, complete original-source identifiers, and end-to-end detector timings are not fully available; the tables therefore establish cache-level reproducibility. No fitted detector-score tail law, empirical mixing verification, or calibrated no-change false-positive rate is claimed here. Those missing measurements delimit the connection between the score-model assumptions and these applications.

B.8.2 UNCERTAINTY IN AGGREGATE AND COMPONENT DIFFERENCES

The per-setting intervals condition on one setting at a time. To assess the family-macro contrasts, we additionally resample the available source components jointly across the entire primary suite. Each draw assigns one multiplicity to a component and reuses it in every setting where that component occurs. We then recompute document means within settings, equally average settings within each family, and equally average the five families. This preserves known cross-setting reuse. The corrected reporting inventory contains 647 metadata-linked components: 643 historical components plus four from the singleton records under the same exact-source/surviving-human-text linkage rule. Missing source identifiers still preclude a claim of full independence.

Table 19: Exploratory family-macro differences: RWCP-R minus comparator, with pointwise paired 95% source-component bootstrap intervals (2,000 draws, seed 121209). Lower WD/MAE and higher F1 are better. Configurations are fixed, linkage is incomplete, and there is no multiplicity correction.

Comparator	RWCP-R minus comparator: paired 95% intervals		
	ΔWD [95% CI]	ΔMAE [95% CI]	ΔF1 [95% CI]
WCP	-0.0959 [-0.1074, -0.0844]	+0.211 [+0.126, +0.293]	-0.057 [-0.073, -0.040]
R: fixed scale	-0.0060 [-0.0099, -0.0021]	-0.052 [-0.071, -0.033]	+0.019 [+0.010, +0.027]
R: uniform weights	-0.0008 [-0.0052, +0.0038]	+0.103 [+0.083, +0.124]	-0.031 [-0.040, -0.023]
R: quadratic approximation	-0.0068 [-0.0108, -0.0033]	+0.074 [+0.052, +0.098]	-0.009 [-0.016, -0.001]

Table 19 sharpens the scope of the benefit. Against WCP, the window-error improvement coexists with higher count MAE and lower boundary F1. Against uniform weights, the WD interval includes zero while the count and F1 intervals favor uniform weights. Against fixed scale, all three intervals favor scale adaptation at the selected configurations. The large-cutoff quadratic approximation trades worse WD for better count and F1. These are conditional comparisons of the saved configurations, not evidence from a newly sampled test corpus or equally retuned alternative models.

B.8.3 PER-SETTING PAIRED RESULTS

The complete standard-WD paired table is Table 20; the preceding appendix tables retain the fuller comparator-specific and legacy-metric views.

Table 20: Primary 25-setting study: standard raw-boundary WD and RWCP-R F1 (tolerance one). Every controlled setting includes 100 records; CoAuthor has 290. Intervals are pointwise paired 95% source-component bootstrap intervals for RWCP-R minus WCP (2,000 draws); negative values favor RWCP-R. Singleton boundary conventions are applied to all methods.

Family	Setting	WCP	RWCP-R		RWCP-R–WCP
		WD	WD	F1 ($h = 1$)	ΔWD [95% CI]
Single	WikiQA	0.284	0.223	0.490	-0.060 [-0.107, -0.013]
Single	News	0.370	0.102	0.790	-0.268 [-0.327, -0.210]
Single	Story	0.534	0.243	0.353	-0.291 [-0.342, -0.238]
CoAuthor	290 sessions	0.488	0.487	0.058	-0.001 [-0.019, 0.015]
Multiple	Claude $K = 1$	0.400	0.288	0.180	-0.112 [-0.156, -0.072]
Multiple	Claude $K = 2$	0.404	0.374	0.140	-0.031 [-0.063, 0.000]
Multiple	Claude $K = 3$	0.425	0.378	0.178	-0.047 [-0.075, -0.019]
Multiple	Claude $K = 5$	0.425	0.429	0.107	0.004 [-0.018, 0.024]
Multiple	Claude $K = 8$	0.430	0.448	0.079	0.018 [0.002, 0.036]
Multiple	GPT $K = 1$	0.434	0.339	0.047	-0.095 [-0.131, -0.060]
Multiple	GPT $K = 2$	0.430	0.408	0.055	-0.022 [-0.047, 0.003]
Multiple	GPT $K = 3$	0.470	0.418	0.071	-0.052 [-0.071, -0.032]
Multiple	GPT $K = 5$	0.475	0.447	0.066	-0.028 [-0.045, -0.012]
Multiple	GPT $K = 8$	0.471	0.464	0.044	-0.006 [-0.019, 0.007]
Attack	WikiQA decoh.	0.340	0.291	0.277	-0.049 [-0.086, -0.011]
Attack	News decoh.	0.407	0.335	0.067	-0.072 [-0.117, -0.030]
Attack	Story decoh.	0.398	0.335	0.067	-0.063 [-0.105, -0.021]
Attack	WikiQA para.	0.266	0.189	0.563	-0.077 [-0.124, -0.034]
Attack	News para.	0.326	0.146	0.648	-0.180 [-0.236, -0.126]
Attack	Story para.	0.428	0.310	0.175	-0.118 [-0.157, -0.079]
Proportion	5%	0.321	0.104	0.037	-0.217 [-0.270, -0.169]
Proportion	10%	0.338	0.172	0.183	-0.166 [-0.219, -0.118]
Proportion	20%	0.375	0.235	0.210	-0.140 [-0.190, -0.097]
Proportion	40%	0.404	0.278	0.203	-0.126 [-0.174, -0.079]
Proportion	80%	0.330	0.270	0.122	-0.060 [-0.097, -0.024]

Table 21: Primary 25-setting legacy final-label window error $\mathrm{WD}_{\mathrm{sup}}$ and RWCP-R raw count error. Best base is the smallest available non-RWCP mean in each setting. Bold marks the smaller legacy error between it and RWCP-R before rounding. These are distinct from raw-boundary WD and matched-boundary F1.

Family	Setting	Best baseline		WCP	RWCP-R	
		Method	$\mathrm{WD}_{\mathrm{sup}}$	$\mathrm{WD}_{\mathrm{sup}}$	$\mathrm{WD}_{\mathrm{sup}}$	CE
Single	WikiQA	VCP	0.200	0.200	**0.184**	0.47
Single	News	WCP	0.123	0.123	**0.097**	0.03
Single	Story	VCP	0.306	0.372	**0.235**	0.34
CoAuthor	290 sessions	VCP	0.480	0.483	**0.478**	10.17
Multiple	Claude $K = 1$	VCP	0.293	0.313	**0.273**	0.58
Multiple	Claude $K = 2$	VCP	**0.319**	0.337	0.360	1.44
Multiple	Claude $K = 3$	VCP	**0.315**	0.339	0.354	2.10
Multiple	Claude $K = 5$	VCP	**0.351**	0.369	0.397	4.26
Multiple	Claude $K = 8$	Voting	**0.348**	0.383	0.409	7.21
Multiple	GPT $K = 1$	WCP	0.423	0.423	**0.342**	0.79
Multiple	GPT $K = 2$	WCP	**0.404**	0.404	0.419	1.51
Multiple	GPT $K = 3$	WCP	0.450	0.450	**0.419**	2.42
Multiple	GPT $K = 5$	WCP	0.438	0.438	**0.427**	4.28
Multiple	GPT $K = 8$	WCP	0.438	0.438	**0.429**	7.22

Continued on next page

Table 21 (continued)

Family	Setting	Best baseline		WCP	RWCP-R	
		Method	$\mathrm{WD}_{\mathrm{sup}}$	$\mathrm{WD}_{\mathrm{sup}}$	$\mathrm{WD}_{\mathrm{sup}}$	CE
Attack	WikiQA decoh.	WCP	0.254	0.254	**0.226**	0.65
Attack	News decoh.	WCP	0.328	0.328	**0.324**	0.60
Attack	Story decoh.	WCP	0.364	0.364	**0.328**	0.59
Attack	WikiQA para.	WCP	0.187	0.187	**0.148**	0.37
Attack	News para.	Voting	**0.107**	0.167	0.141	0.12
Attack	Story para.	WCP	0.336	0.336	**0.303**	0.52
Proportion	5%	WCP	0.213	0.213	**0.111**	0.79
Proportion	10%	WCP	0.235	0.235	**0.185**	0.49
Proportion	20%	WCP	**0.228**	0.228	0.243	0.48
Proportion	40%	WCP	0.289	0.289	**0.277**	0.52
Proportion	80%	WCP	0.275	0.275	**0.271**	0.72

C Theoretical Proofs

C.1 Technical Preliminaries and Population Geometry

The master collection is $\mathcal{A}_N = \{[s,e] : 1 \le s \le e \le N\}$; subintervals arising in any recursive call therefore belong to it. For $I = [s,e]$, the admissible set is

$$\mathcal{B}(I) = \{b \in \{s, \dots, e-1\} : S^w_{s:b} \ge m_N,\ S^w_{b+1:e} \ge m_N\}.$$

Set $T(I) = -\infty$ if this set is empty. Draw two independent uniform coordinates on $(0, W_N]$, map each to $\min\{i : W(i) \ge u\}$, and sort the resulting endpoints. Draw the candidate collection once, independently of the scores conditional on deterministic weights; recursive calls retain only intervals fully contained in their search range. Ties are resolved deterministically by endpoints and then split index. The uniform events cover the full master collection, not only the sampled intervals.

This appendix collects the deterministic notation and population-risk arguments underlying the profile margin in Section 3. It first records merge identities, then establishes the behavior of Huber minimizers, and finally proves the single-change population result.

For an interval $A = [a,b]$, write $S^w_A = S^w_{a:b}$ and

$$\widehat{L}^\star_A = \inf_{\theta\in\Theta} \widehat{\mathcal{L}}_A(\theta), \qquad L^\star_A = \inf_{\theta\in\Theta} \mathcal{L}_A(\theta).$$

For adjacent intervals A, B, empirical and population merge costs are

$$\widehat{D}(A,B) = \widehat{L}^\star_{A\cup B} - \widehat{L}^\star_A - \widehat{L}^\star_B,$$
$$D^\star(A,B) = L^\star_{A\cup B} - L^\star_A - L^\star_B.$$

Both are nonnegative because a common parameter is more restrictive than separate parameters.

Lemma C.1 (Profile-gain difference identity)**.** *Suppose $s \le b < \tau < e$ and set*

$$A = [s,b], \qquad B = [b+1,\tau], \qquad C = [\tau+1,e].$$

Then

$$\Gamma^\star_{s,e}(\tau) - \Gamma^\star_{s,e}(b) = 2\{D^\star(B,C) - D^\star(A,B)\},$$
$$\widehat{\Gamma}_{s,e}(\tau) - \widehat{\Gamma}_{s,e}(b) = 2\{\widehat{D}(B,C) - \widehat{D}(A,B)\}.$$

If A and B belong to the same true segment, then $D^\star(A,B) = 0$.

Proof. By definition,

$$\frac{1}{2}\Gamma^\star_{s,e}(\tau) = L^\star_{A\cup B\cup C} - L^\star_{A\cup B} - L^\star_C$$

and

$$\frac{1}{2}\Gamma^\star_{s,e}(b) = L^\star_{A\cup B\cup C} - L^\star_A - L^\star_{B\cup C}.$$

Subtracting cancels the parent term and gives

$$\frac{1}{2}\{\Gamma^\star_{s,e}(\tau) - \Gamma^\star_{s,e}(b)\} = L^\star_A + L^\star_{B\cup C} - L^\star_{A\cup B} - L^\star_C.$$

Add and subtract $L^\star_B$ to obtain

$$\{L^\star_{B\cup C} - L^\star_B - L^\star_C\} - \{L^\star_{A\cup B} - L^\star_A - L^\star_B\},$$

which is $D^\star(B, C) - D^\star(A, B)$. The empirical identity is identical with hats. If A and B share minimizer θ_-, then

$$L^\star_{A\cup B} \le \mathcal{L}_A(\theta_-) + \mathcal{L}_B(\theta_-) = L^\star_A + L^\star_B.$$

The reverse inequality follows from separate minimization, so equality holds and $D^\star(A, B) = 0$. □

Lemma C.2 (Harmonic-mass inequalities)**.** *For $x, y > 0$,*

$$\frac{1}{2}\min(x, y) \le \frac{xy}{x+y} \le \min(x, y). \tag{22}$$

Moreover, for all $a, b, c \ge 0$,

$$(a - b - c)_+^2 \ge \frac{1}{2}a^2 - 2b^2 - 2c^2. \tag{23}$$

Proof. Assume $x \le y$. Then $x + y \le 2y$ gives $xy/(x + y) \ge x/2$, while $x + y \ge y$ gives $xy/(x+y) \le x$. For the second inequality, if $a \le b+c$, the right side is at most $a^2/2 - (b+c)^2 \le 0$. If $a > b + c$, use $(a - d)^2 = a^2 - 2ad + d^2 \ge a^2/2 - d^2$ with $d = b + c$, followed by $(b + c)^2 \le 2b^2 + 2c^2$. □

C.1.1 Population geometry proofs

Symmetric-location proposition.

Proof. Fix i in segment j and write $a_i = \sigma_i/s_i > 0$. For $u = \theta - \theta_j$,

$$Q_i(\theta_j + u) = \mathbb{E}\rho_c(a_i\xi_i - u/s_i).$$

The function is convex in u. At $u = 0$, an interior subgradient is

$$-\frac{1}{s_i}\mathbb{E}\psi_c(a_i\xi_i).$$

The Huber score is odd and the law of ξ_i is symmetric, hence $\mathbb{E}\psi_c(a_i\xi_i) = 0$. Therefore zero belongs to the subdifferential and $u = 0$ minimizes the convex risk. If $\mathbb{E}|\xi_i| < \infty$ and $\mathbb{E}\xi_i = 0$, then

$$\mathbb{E}Y_i = \theta_j + \sigma_i\mathbb{E}\xi_i = \theta_j.$$

This last conclusion concerns the ordinary mean and uses integrability; it is not implied by Huber centering alone. □

Existence and uniqueness of interval minimizers.

Lemma C.3 (Existence)**.** *For every nonempty interval A, the empirical loss $\widehat{\mathcal{L}}_A$ and population loss $\mathcal{L}_A$ attain minima on compact Θ.*

Proof. The Huber loss is continuous. Therefore $\widehat{\mathcal{L}}_A$, a finite sum of continuous functions, is continuous on compact Θ and attains its minimum. Since ρ_c is nonnegative and Lipschitz with constant c, dominated convergence on the compact parameter set gives continuity of each Q_i whenever $\mathbb{E}\rho_c(\sqrt{w_i}(Y_i - \theta_0)) < \infty$ for one $\theta_0 \in \Theta$. Thus $\mathcal{L}_A$ is continuous and also attains its minimum. □

Lemma C.4 (Uniqueness under curvature)**.** *If A is contained in one segment and Assumption 3.2 holds, then $\theta^\star_A$ is unique and equals that segment's θ_j.*

Proof. Summing (7) over $i \in A$ yields

$$\mathcal{L}_A(\theta) - \mathcal{L}_A(\theta_j) \ge \frac{m_0}{2}S^w_A(\theta - \theta_j)^2$$

for $|\theta - \theta_j| \le r_0$. If a distinct global minimizer existed, convexity would force the aggregate risk to be constant on the entire line segment joining it to θ_j. That segment contains points arbitrarily close to θ_j, contradicting the strictly positive local quadratic lower bound above. Hence θ_j is the unique interval minimizer. □

Mixed-segment excess risk.

Proof. Let $x = S_A^w$, $y = S_B^w$, $a = \theta_A$, and $b = \theta_B$. By curvature,

$$\mathcal{L}_A(\theta) - L_A^\star \geq \frac{m_0 x}{2}(\theta - a)^2,$$
$$\mathcal{L}_B(\theta) - L_B^\star \geq \frac{m_0 y}{2}(\theta - b)^2$$

for every θ on the segment joining a and b. A minimizer of the sum lies between a and b because the first derivative is nonpositive to the left of both minimizers and nonnegative to the right. Therefore

$$\begin{aligned} D^\star(A, B) &= \inf_\theta \{\mathcal{L}_A(\theta) - L_A^\star + \mathcal{L}_B(\theta) - L_B^\star\} \\ &\geq \frac{m_0}{2} \inf_\theta \{x(\theta - a)^2 + y(\theta - b)^2\}. \end{aligned}$$

The quadratic minimizer is $(xa + yb)/(x + y)$ and the minimum is

$$\frac{xy}{x + y}(a - b)^2.$$

This proves (9). The upper curvature bound evaluated at the same weighted average gives (10). □

Single-change population margin.

Proof. Consider $b < \tau$ and use the intervals A, B, C from Lemma C.1. Since A and B belong to the left segment, $D^\star(A, B) = 0$. Therefore

$$\Gamma_{s,e}^\star(\tau) - \Gamma_{s,e}^\star(b) = 2D^\star(B, C).$$

Lemma 3.7 gives

$$2D^\star(B, C) \geq m_0 \kappa^2 \frac{B_w C_w}{B_w + C_w},$$

which proves (11). The case $b > \tau$ is symmetric after exchanging left and right. The difference is nonnegative for every b and is strictly positive when the candidate differs from τ, so the true split is the unique maximizer among admissible splits. Finally, if $B_w \leq C_w$, Lemma C.2 yields

$$\frac{B_w C_w}{B_w + C_w} \geq \frac{1}{2} B_w = \frac{1}{2} d_w(b, \tau).$$

The other cases are identical. □

Quadratic-loss reduction.

Proof. Under $\rho_\infty(u) = u^2/2$,

$$\widehat{\mathcal{L}}_{a:b}(\theta) = \frac{1}{2} \sum_{i=a}^{b} w_i (Y_i - \theta)^2$$

and the minimizer is $\bar{Y}_{a:b}^w$. The weighted analysis-of-variance identity gives

$$\begin{aligned} &\sum_{i=s}^{e} w_i (Y_i - \bar{Y}_{s:e}^w)^2 - \sum_{i=s}^{b} w_i (Y_i - \bar{Y}_{s:b}^w)^2 - \sum_{i=b+1}^{e} w_i (Y_i - \bar{Y}_{b+1:e}^w)^2 \\ &\qquad = \frac{S_{s:b}^w S_{b+1:e}^w}{S_{s:e}^w} (\bar{Y}_{s:b}^w - \bar{Y}_{b+1:e}^w)^2. \end{aligned}$$

Multiplication by the factor two in (3) cancels the one-half in the loss, proving (4). □

C.2 Stochastic Control and Exact-Recovery Proof

This appendix supplies the stochastic part of the recovery proof. We state a valid dependence-aware concentration bound, transfer it to interval fits and merge costs, and combine these bounds with weighted interval isolation and boundary-fragment control to prove exact recovery. For a homogeneous interval A with true robust location θ_A, write

$$U_A = \sum_{i \in A} \sqrt{w_i}\, \psi_c(\sqrt{w_i}(Y_i - \theta_A)) .$$

Lemma C.5 (Uniform Huber-score concentration)**.** *Under Assumptions 3.1, 3.3, and 3.4, with probability at least* $1 - \delta/8$*,*

$$|U_A| \le C_{\mathrm{dep}} \left\{ \sqrt{S_A^w x_N(\delta)} + q_N x_N(\delta) \right\}$$

simultaneously for every homogeneous $A \in \mathcal{A}_N$*, without a minimum-mass restriction. Moreover, whenever* $S_A^w \ge m_{\mathrm{curv}}$*,*

$$|U_A| \le C'_{\mathrm{dep}} \sqrt{S_A^w \Lambda_N(\delta)}. \tag{24}$$

Lemma C.6 (Uniform empirical curvature)**.** *Under Assumptions 3.3–3.5, with probability at least* $1 - \delta/8$*,*

$$\widehat{\mathcal{L}}_A(\theta) - \widehat{\mathcal{L}}_A(\theta_A) \ge -U_A(\theta - \theta_A) + \frac{p_0}{8} S_A^w (\theta - \theta_A)^2 \tag{25}$$

for every homogeneous $A \in \mathcal{A}_N$ *with* $S_A^w \ge m_{\mathrm{curv}}$ *and every* $|\theta - \theta_A| \le r_0/2$*.*

We first establish a blocking inequality used for Huber scores and empirical curvature indicators. Constants in this appendix are explicit functions of the mixing parameters but are denoted by C to avoid obscuring the rates.

Bernstein input.

Lemma C.7 (Dependence-aware Bernstein inequality)**.** *Let* $V_1, \dots, V_n$ *be centered, coordinate-wise measurable variables with* $|V_i| \le B$*. Under Assumption 3.4, for every* $x \ge 1$*,*

$$\mathbb{P}\left(\left| \sum_{i=1}^n V_i \right| > C_{\mathrm{dep}} B \{ \sqrt{nx} + q_N x \} \right) \le 2e^{-x}.$$

For independent or fixed-order dependent sequences, $q_N = O(1)$*. For geometrically beta-mixing bounded sequences in the sense of* (8)*, one may take the conservative* $q_N = O\{\log^2(eN)\}$ *specialization of the published strong-mixing Bernstein inequality, with constants depending on* (C_β, c_β)*.*

Proof. The first claim is Assumption 3.4. The independent and fixed-order cases follow from ordinary Bernstein blocking. Geometric beta mixing implies geometric alpha mixing; specializing the bounded-variable inequality of Merlevède et al. (2009) gives a linear term of order $x \log^2(en)$. We retain that factor through q_N and do not invoke a pairwise-covariance-to-cumulant shortcut. □

Uniform Huber-score concentration.

Proof. For a homogeneous interval $A = [a, b]$, the summands in U_A are centered by (1) and bounded by

$$B_Z = c\sqrt{w_{\max}}.$$

Apply Lemma C.7 to the subsequence indexed by A, noting that restriction to a contiguous interval preserves the mixing bound. Since $|A| \le S_A^w / w_{\min}$, Lemma C.7 gives

$$\begin{aligned} |U_A| &\le C B_Z \left[\sqrt{\frac{S_A^w}{w_{\min}} x} + q_N x \right] \\ &\le C_{\mathrm{dep}} \left[\sqrt{S_A^w x} + q_N x \right]. \end{aligned}$$

Use $x = x_N(\delta)$ and take a union bound over at most L_N intervals. Because $2L_N e^{-x_N(\delta)} \le \delta/32$, the simultaneous failure probability is at most $\delta/8$. Finally, $S_A^w \ge m_{\mathrm{curv}} \gtrsim q_N^2 x_N$ turns the displayed bound into inequality (24). □

Uniform empirical curvature. For any interval A, define the empirical Hessian where it exists,

$$H_A(\theta) = \sum_{i \in A} w_i \mathbf{1}\left\{|\sqrt{w_i}(Y_i - \theta)| < c\right\}. \tag{26}$$

Lemma C.8 (Uniform inlier mass with adjacent references)**.** *Under Assumptions 3.3, 3.4, and 3.5, with probability at least $1 - \delta/8$, the following holds uniformly. Let $A \in \mathcal{A}_N$ be homogeneous or intersect at most two adjacent true segments. Let θ_0 either be a true location represented in A or, when A is homogeneous, the location of a true segment immediately adjacent to the segment represented in A. Suppose $S_A^w \geq m_{\mathrm{curv}}$. Then*

$$H_A(\theta) \geq \frac{p_0}{2} S_A^w \tag{27}$$

for every $|\theta - \theta_0| \leq r_0/2$, provided C_{curv} is sufficiently large.

Proof. For each eligible pair (A, θ_0), construct a deterministic grid $\mathcal{G}$ on $[-r_0/2, r_0/2]$ with mesh

$$\eta = \frac{c}{8\sqrt{w_{\max}}}.$$

For $u \in \mathcal{G}$, define

$$I_i(u) = \mathbf{1}\left\{|\sqrt{w_i}\{Y_i - (\theta_0 + u)\}| \leq \frac{3c}{4}\right\}.$$

For every eligible pair and every $i \in A$, the locations θ_0 and $\theta_{z(i)}$ are identical or belong to adjacent true segments. Hence

$$|\theta_0 + u - \theta_{z(i)}| \leq r_0/2 + \kappa_{\max} \leq 3r_0/4 < r_0,$$

where $\kappa_{\max} = 0$ in the no-change case. Assumption 3.5 therefore gives $\mathbb{E}I_i(u) \geq p_0$; the larger $3c/4$ window contains the $c/2$ event used in that assumption. The centered variables

$$V_i(u) = w_i\{I_i(u) - \mathbb{E}I_i(u)\}$$

are bounded by $w_{\max}$. Apply Lemma C.7 at $x = x_N(\delta) + \log(16|\mathcal{G}|)$ and take a union bound over the at most three eligible reference locations per interval and the fixed grid. The additive constant in x is absorbed into C, giving simultaneously

$$\left|\sum_{i \in A} V_i(u)\right| \leq C\left\{\sqrt{S_A^w x_N} + q_N x_N\right\}.$$

The condition $S_A^w \geq m_{\mathrm{curv}}$ and a sufficiently large C_{curv} make the right side at most $(p_0/2)S_A^w$. Hence

$$\sum_{i \in A} w_i I_i(u) \geq \frac{p_0}{2} S_A^w.$$

For arbitrary $|v| \leq r_0/2$, choose $u \in \mathcal{G}$ with $|u - v| \leq \eta$. If $I_i(u) = 1$, then

$$|\sqrt{w_i}\{Y_i - (\theta_0 + v)\}| \leq \frac{3c}{4} + \sqrt{w_i}|u - v| < c.$$

Thus the grid inlier set is contained in the Hessian inlier set at v, which proves (27) for every eligible interval–reference pair. □

Proof of Lemma C.6. Fix a homogeneous A and write $d = \theta - \theta_A$. The fundamental theorem for convex functions gives

$$\widehat{\mathcal{L}}_A(\theta) - \widehat{\mathcal{L}}_A(\theta_A) = -U_A d + \int_0^d (d - u) H_A(\theta_A + u)\,\mathrm{d}u$$

when $d \geq 0$, with the analogous oriented integral for $d < 0$. Lemma C.8 lower bounds H_A by $(p_0/2)S_A^w$ along the path. Hence the integral is at least

$$\frac{p_0}{4} S_A^w d^2.$$

The claimed coefficient $p_0/8$ is weaker and therefore holds uniformly. □

C.2.1 Uniform Huber fits and empirical merge costs

Lemma C.9 (Homogeneous interval fit)**.** *On the intersection of the score and curvature events, every homogeneous $A \in \mathcal{A}_N$ with $S_A^w \geq m_{\mathrm{curv}}$ satisfies*

$$|\widehat{\theta}_A - \theta_A| \leq C_\theta \sqrt{\frac{\Lambda_N(\delta)}{S_A^w}} \leq r_0/4 \tag{28}$$

and

$$0 \leq \widehat{\mathcal{L}}_A(\theta_A) - \widehat{\mathcal{L}}_A(\widehat{\theta}_A) \leq C_L \Lambda_N(\delta). \tag{29}$$

The final $r_0/4$ bound is enforced by the constant in m_{curv}.

Lemma 3.9 is stated in Section 3; the following arguments prove its three cases.

Homogeneous interval fit.

Proof. Work on the score and curvature events. Let $S = S_A^w$, $U = U_A$, $\theta_0 = \theta_A$, and $d = \theta - \theta_0$. By (25),

$$\widehat{\mathcal{L}}_A(\theta) - \widehat{\mathcal{L}}_A(\theta_0) \geq -|U||d| + \alpha S d^2, \qquad \alpha = p_0/8. \tag{30}$$

Lemma C.5 and $S \geq m_{\mathrm{curv}} = C_{\mathrm{curv}} \Lambda_N$ give

$$\frac{|U|}{S} \leq C \sqrt{\frac{\Lambda_N}{S}} \leq \alpha r_0/4$$

after increasing C_{curv}. Evaluating (30) at $|d| = r_0/2$ then gives a strictly positive loss difference. Convexity rules out a minimizer outside these two boundary points, so the full path to every minimizer lies in the curvature region and

$$|\widehat{\theta}_A - \theta_0| \leq \frac{|U|}{\alpha S} \leq C \sqrt{\frac{\Lambda_N}{S}} \leq r_0/4.$$

Because $\mathrm{dist}(\theta_0, \partial\Theta) \geq r_0$, the fitted location is interior and has zero subgradient. This proves (28). For the loss improvement, minimization and (30) imply

$$\begin{aligned} 0 &\leq \widehat{\mathcal{L}}_A(\theta_0) - \widehat{\mathcal{L}}_A(\widehat{\theta}_A) \\ &\leq \sup_{d \in \mathbb{R}} \{|U||d| - \alpha S d^2\} = \frac{U^2}{4\alpha S} \leq C_L \Lambda_N, \end{aligned}$$

which is (29). □

Empirical strong convexity around fitted locations.

Lemma C.10 (Local empirical strong convexity)**.** *On the event of Lemma C.8, let $A \in \mathcal{A}_N$ be homogeneous with $S_A^w \geq m_{\mathrm{curv}}$, and suppose $\widehat{\theta}_A$ and θ both lie within $r_0/2$ of the true location. Then*

$$\widehat{\mathcal{L}}_A(\theta) - \widehat{\mathcal{L}}_A(\widehat{\theta}_A) \geq \frac{p_0}{4} S_A^w (\theta - \widehat{\theta}_A)^2.$$

Furthermore, for all θ, θ',

$$\widehat{\mathcal{L}}_A(\theta) \leq \widehat{\mathcal{L}}_A(\theta') + g_A(\theta')(\theta - \theta') + \frac{1}{2} S_A^w (\theta - \theta')^2, \tag{31}$$

where $g_A(\theta')$ is any derivative at a differentiability point or subgradient otherwise.

Proof. Integrating the Hessian lower bound between $\widehat{\theta}_A$ and θ gives strong convexity with coefficient $(p_0/2)S_A^w$ in the conventional definition, hence the factor $p_0/4$ in the function inequality. For smoothness, the Huber score is one-Lipschitz, so each map $\theta \mapsto \rho_c(\sqrt{w_i}(Y_i - \theta))$ has derivative Lipschitz constant at most w_i. Summing gives Lipschitz gradient constant S_A^w, which yields (31). □

Empirical merge-cost bounds.

Proof. We work on the uniform score, estimator, and curvature events.

Same-location upper bound. Suppose A and B share true location θ_0. Since the merged optimum is no worse than evaluating the common parameter θ_0,

$$\begin{aligned}\widehat{D}(A,B) &= \widehat{L}^\star_{A\cup B} - \widehat{L}^\star_A - \widehat{L}^\star_B \\ &\le \widehat{\mathcal{L}}_A(\theta_0) + \widehat{\mathcal{L}}_B(\theta_0) - \widehat{L}^\star_A - \widehat{L}^\star_B \\ &= \{\widehat{\mathcal{L}}_A(\theta_0) - \widehat{\mathcal{L}}_A(\widehat{\theta}_A)\} + \{\widehat{\mathcal{L}}_B(\theta_0) - \widehat{\mathcal{L}}_B(\widehat{\theta}_B)\} \\ &\le 2C_L\Lambda_N.\end{aligned}$$

Nonnegativity follows from nesting. This proves (12).

Different-location lower bound. Let true locations be a and b, jump $\kappa = |a-b|$, masses $x = S^w_A$ and $y = S^w_B$, and estimates $\widehat{a}, \widehat{b}$. Lemma C.9 gives $|\widehat{a}-a|, |\widehat{b}-b| \le r_0/4$; combined with $\kappa \le r_0/4$, every point between the two estimates lies within $r_0/2$ of each relevant true location. Lemma C.10 therefore applies and gives

$$\begin{aligned}\widehat{\mathcal{L}}_A(\theta) - \widehat{\mathcal{L}}_A(\widehat{a}) &\ge \frac{p_0 x}{4}(\theta-\widehat{a})^2, \\ \widehat{\mathcal{L}}_B(\theta) - \widehat{\mathcal{L}}_B(\widehat{b}) &\ge \frac{p_0 y}{4}(\theta-\widehat{b})^2.\end{aligned}$$

The minimizer of the merged convex loss lies between $\widehat{a}$ and $\widehat{b}$. Therefore

$$\widehat{D}(A,B) \ge \frac{p_0}{4}\frac{xy}{x+y}(\widehat{a}-\widehat{b})^2.$$

Estimator localization gives

$$|\widehat{a}-\widehat{b}| \ge \kappa - C_\theta\sqrt{\Lambda_N/x} - C_\theta\sqrt{\Lambda_N/y}.$$

Apply (23) and multiply by $xy/(x+y)$. Since

$$\frac{xy}{x+y}\frac{1}{x} \le 1, \qquad \frac{xy}{x+y}\frac{1}{y} \le 1,$$

we obtain

$$\widehat{D}(A,B) \ge \frac{p_0}{8}\kappa^2\frac{xy}{x+y} - C\Lambda_N,$$

which is (13).

Different-location upper bound. The interior margin and Lemma C.9 ensure that both separate minimizers are interior, so their subgradients vanish. Global smoothness therefore gives

$$\begin{aligned}\widehat{\mathcal{L}}_A(\theta) - \widehat{\mathcal{L}}_A(\widehat{a}) &\le \frac{x}{2}(\theta-\widehat{a})^2, \\ \widehat{\mathcal{L}}_B(\theta) - \widehat{\mathcal{L}}_B(\widehat{b}) &\le \frac{y}{2}(\theta-\widehat{b})^2.\end{aligned}$$

Evaluate the merged objective at $\theta = (x\widehat{a} + y\widehat{b})/(x+y)$ to get

$$\widehat{D}(A,B) \le \frac{1}{2}\frac{xy}{x+y}(\widehat{a}-\widehat{b})^2.$$

Using $(u+v+w)^2 \le 3(u^2+v^2+w^2)$ and the estimator bounds gives

$$\widehat{D}(A,B) \le C\kappa^2\frac{xy}{x+y} + C\Lambda_N,$$

which proves (14). □

C.2.2 WEIGHTED INTERVAL ISOLATION

For the proof of Lemma 3.10, for each change τ_j, define weighted endpoint neighborhoods

$$\mathcal{L}_j = \left\{ s \le \tau_j : \frac{\Delta_w}{6} \le S^w_{s:\tau_j} \le \frac{\Delta_w}{3} \right\},$$
$$\mathcal{R}_j = \left\{ e > \tau_j : \frac{\Delta_w}{6} \le S^w_{\tau_j+1:e} \le \frac{\Delta_w}{3} \right\}.$$

Proof of Lemma 3.10. Fix change τ_j. The segment to its left has mass at least Δ_w. Move left from τ_j until the accumulated mass first reaches $\Delta_w/6$. Because one atom is at most $\Delta_w/24$, the attained mass is at most $5\Delta_w/24 < \Delta_w/3$. Continuing until just before mass exceeds $\Delta_w/3$ shows that the set of weighted coordinates corresponding to valid left endpoints has length at least

$$\frac{\Delta_w}{3} - \frac{\Delta_w}{6} - 2\max_i w_i \ge \frac{\Delta_w}{12}.$$

The same argument holds on the right.

Under weighted-coordinate sampling, the probability that one ordered pair of endpoints falls in the left and right neighborhoods is at least

$$p_j \ge 2\left(\frac{\Delta_w/12}{W_N}\right)^2 = \frac{\Delta_w^2}{72W_N^2},$$

where the factor two accounts for the two orders before sorting. Hence the probability that none of M intervals isolates τ_j is at most

$$(1-p_j)^M \le \exp(-Mp_j).$$

With (15), this is at most $\delta/[4(K \vee 1)]$. A union bound over K changes gives failure probability at most $\delta/4$ and proves simultaneous isolation; for $K = 0$ the statement is vacuous. The endpoint construction keeps s and e inside the adjacent segments, so the interval contains no other change. The mass bounds follow directly from the definitions of $\mathcal{L}_j$ and $\mathcal{R}_j$. □

C.2.3 BOUNDARY FRAGMENTS AND EXACT RECOVERY

The constants can be selected in dependency order. First choose C_{curv} large enough for the score and inlier-mass bounds under the fixed $c, p_0, r_0, w_{\min}, w_{\max}, C_{\mathrm{dep}}$. Fix a jump range and $c_{m,1} \ge C_{\mathrm{curv}}\kappa^2_{\max}$. Next choose $c_{\mathrm{cert}} > \max\{c_{m,1}, 2(C_0+C_1)/c_0\}$, then $C_r > \max\{2C_0, C_{\mathrm{frag}}(1 + c_{\mathrm{cert}})\}$. Choose $C_{\mathrm{loc}} > \max\{C_{\mathrm{curv}}\kappa^2_{\max}, 2(C_0+C_1)/c_0\}$, and $C_g > C_{\mathrm{loc}}$. Finally enlarge C_{snr} to make isolating gains exceed $C_r\Lambda_N$, localization at most $\Delta_w/12$, and the guard/side-mass intervals nonempty, including atom overshoot. These constants depend on the stated fixed model parameters, not on the document-specific score realization.

We now prove Theorem 3.11. The proof uses an additional deterministic consequence of bounded Huber influence: a true segment fragment with small reliability mass cannot by itself create an arbitrarily large profile gain, even if the raw scores in that fragment are extreme.

A small-fragment gain bound.

Lemma C.11 (Score transfer between adjacent locations)**.** *Let A lie in a segment with location θ_1, and let θ_2 be an adjacent location with $\kappa = |\theta_1 - \theta_2|$. Define*

$$U_A(\theta) = \sum_{i\in A} \sqrt{w_i}\,\psi_c\{\sqrt{w_i}(Y_i - \theta)\}.$$

Then

$$|U_A(\theta_2)| \le |U_A(\theta_1)| + \kappa S^w_A.$$

Proof. The Huber score is one-Lipschitz. Hence, term by term,

$$\begin{aligned}&|\sqrt{w_i}\psi_c\{\sqrt{w_i}(Y_i-\theta_2)\} - \sqrt{w_i}\psi_c\{\sqrt{w_i}(Y_i-\theta_1)\}| \\ &\quad\le w_i|\theta_2-\theta_1| = w_i\kappa.\end{aligned}$$

Summing and applying the triangle inequality proves the claim. □

Lemma C.12 (Fit improvement relative to a reference location)**.** *Let A intersect at most two adjacent true segments. Let θ_0 either be represented in A or, when A is homogeneous, be the location of a true segment immediately adjacent to the segment represented in A. Define $r_A = \sum_{i\in A:\,\theta_{z(i)}\neq\theta_0} w_i$ and $\kappa = \max_{i\in A}|\theta_{z(i)} - \theta_0|$. Suppose $S_A^w \geq m_{\mathrm{curv}}$. On the uniform score and empirical-curvature event,*

$$0 \leq \widehat{\mathcal{L}}_A(\theta_0) - \inf_{\theta\in\Theta} \widehat{\mathcal{L}}_A(\theta) \leq C\left\{\Lambda_N + \kappa^2 \frac{r_A^2}{S_A^w}\right\}.$$

Proof. Decompose $A = A_0 \cup A_1$, where A_0 contains the observations whose location is θ_0 and $A_1 = A \setminus A_0$. By eligibility, all observations in a nonempty A_1 share one location adjacent to θ_0, and $S_{A_1}^w = r_A$; either set may be empty. The score at θ_0 is

$$U_A(\theta_0) = U_{A_0}(\theta_0) + U_{A_1}(\theta_0).$$

Uniform score concentration on the nonempty homogeneous fragments and Lemma C.11 give

$$|U_A(\theta_0)| \leq C\sqrt{S_A^w \Lambda_N} + \kappa r_A. \tag{32}$$

By the represented-or-adjacent reference clause of Lemma C.8, the empirical Hessian on A is at least αS_A^w for every $|\theta - \theta_0| \leq r_0/2$, with fixed $\alpha = p_0/2$. Because $S_A^w \geq C_{\mathrm{curv}}\Lambda_N$ and $\kappa \leq p_0 r_0/16$, (32) is at most $\alpha S_A^w r_0/4$ after choosing the constants. Evaluating the quadratic lower bound at $\theta_0 \pm r_0/2$ gives a positive loss difference; convexity therefore places every mixed-interval minimizer inside that interval. Only after this containment step do we integrate the empirical curvature along the path, obtaining

$$\widehat{\mathcal{L}}_A(\theta_0) - \inf_{\theta} \widehat{\mathcal{L}}_A(\theta) \leq C\frac{U_A^2(\theta_0)}{S_A^w}.$$

Finally, (32) and $(a+b)^2 \leq 2a^2 + 2b^2$ give

$$\frac{U_A^2(\theta_0)}{S_A^w} \leq C\left\{\Lambda_N + \kappa^2 \frac{r_A^2}{S_A^w}\right\}.$$

□

Lemma C.13 (Small boundary fragment)**.** *Let $I = [s, e]$ contain exactly one change with jump κ, and let*

$$r(I) = \min\{S_{s:\tau}^w, S_{\tau+1:e}^w\}.$$

If every admissible child has mass at least $m_N \geq m_{\mathrm{curv}}$, then on the uniform event,

$$\max_{b\in\mathcal{B}(I)} \widehat{\Gamma}_{s,e}(b) \leq C_{\mathrm{frag}}\{\Lambda_N + \kappa^2 r(I)\}. \tag{33}$$

Proof. Assume without loss of generality that the left true segment is the smaller one, with location θ_- and mass $r = r(I)$, while the right segment has location θ_+. For an admissible candidate b, let $L = [s, b]$ and $R = [b+1, e]$. Since the one-location parent fit is no larger than evaluation at θ_+,

$$\begin{aligned}\frac{1}{2}\widehat{\Gamma}_{s,e}(b) &= \widehat{L}_{L\cup R}^\star - \widehat{L}_L^\star - \widehat{L}_R^\star \\ &\leq \{\widehat{\mathcal{L}}_L(\theta_+) - \widehat{L}_L^\star\} + \{\widehat{\mathcal{L}}_R(\theta_+) - \widehat{L}_R^\star\}.\end{aligned}$$

Let r_L, r_R be the left-segment masses contained in L, R, so $r_L + r_R = r$. Lemma C.12 gives

$$\begin{aligned}\widehat{\Gamma}_{s,e}(b) &\leq C\left\{\Lambda_N + \kappa^2\frac{r_L^2}{S_L^w} + \Lambda_N + \kappa^2\frac{r_R^2}{S_R^w}\right\} \\ &\leq C\{\Lambda_N + \kappa^2(r_L + r_R)\},\end{aligned}$$

where $r_L^2/S_L^w \leq r_L$ and similarly for R. This is (33). □

No false positives and isolated signal.

Lemma C.14 (No false positive)**.** *If $I = [s, e]$ contains no true change, then on the event of Lemma 3.9,*

$$T(I) \le 2C_0\Lambda_N.$$

Thus $T(I) \le r_N$ whenever $C_r > 2C_0$.

Proof. For every admissible b, the two child intervals belong to the same segment. By (3),

$$\widehat{\Gamma}_{s,e}(b) = 2\widehat{D}([s, b], [b + 1, e]).$$

Apply (12) and maximize over b. □

Lemma C.15 (Signal on an isolating interval)**.** *Let $I = [s, e]$ isolate change τ_j and satisfy* (16)*. If $m_N \le \Delta_w/12$, then τ_j is admissible and*

$$\widehat{\Gamma}_{s,e}(\tau_j) \ge c_{\mathrm{sig}}\kappa_j^2\Delta_w - C_{\mathrm{sig}}\Lambda_N.$$

Consequently $T(I) > r_N$ under the signal condition with a sufficiently large C_{snr}.

Proof. The two true-side masses are at least $\Delta_w/6 > m_N$, so the split is admissible. Moreover,

$$\widehat{\Gamma}_{s,e}(\tau_j) = 2\widehat{D}([s, \tau_j], [\tau_j + 1, e]).$$

Lemma 3.9 and Lemma C.2 yield

$$\begin{aligned}\widehat{\Gamma}_{s,e}(\tau_j) &\ge 2c_0\kappa_j^2 \frac{S^w_{s:\tau_j} S^w_{\tau_j+1:e}}{S^w_{s:e}} - 2C_1\Lambda_N \\ &\ge c_{\mathrm{sig}}\kappa_j^2\Delta_w - C_{\mathrm{sig}}\Lambda_N.\end{aligned}$$

The signal condition makes this larger than $C_r\Lambda_N$. □

The selected interval contains one well-supported change.

Lemma C.16 (Narrowest interval contains one change)**.** *On the no-false-positive and isolation events, every recursive search range containing an unresolved change has a nonempty over-threshold set. The selected narrowest interval contains exactly one unresolved true change.*

Proof. Let τ_j be unresolved in the current search range. The induction argument in Lemma C.21 below shows that one of its isolating intervals remains fully inside the range. Lemma C.15 places that interval in the over-threshold set, so the set is nonempty. Its weighted length is at most $2\Delta_w/3$. Hence the selected interval has weighted length at most $2\Delta_w/3$.

An interval containing no change cannot be over threshold by Lemma C.14. An interval containing two or more consecutive changes must contain the complete segment between two adjacent changes, whose mass is at least Δ_w. Such an interval has weighted length at least $\Delta_w > 2\Delta_w/3$. Therefore the selected interval contains exactly one change. □

Lemma C.17 (True-side mass certification)**.** *Choose $C_r > C_{\mathrm{frag}}(1 + c_{\mathrm{cert}})$ for a constant $c_{\mathrm{cert}} > \max\{c_{m,1}, 2(C_0 + C_1)/c_0\}$. Every selected one-change interval I with jump κ_j satisfies*

$$r(I) \ge c_{\mathrm{cert}} \frac{\Lambda_N}{\kappa_j^2} \ge m_N.$$

In particular, the true split is admissible.

Proof. Selection implies $T(I) > C_r\Lambda_N$. Lemma C.13 gives

$$C_r\Lambda_N < T(I) \le C_{\mathrm{frag}}\{\Lambda_N + \kappa_j^2 r(I)\}.$$

Rearranging yields

$$r(I) > \left(\frac{C_r}{C_{\mathrm{frag}}} - 1\right)\frac{\Lambda_N}{\kappa_j^2} \ge c_{\mathrm{cert}}\frac{\Lambda_N}{\kappa_j^2}.$$

Because $\kappa_j \le \kappa_{\max}$ and $m_N \le c_{m,1}\Lambda_N/\kappa_{\max}^2$, the final quantity is at least m_N when $c_{\mathrm{cert}} > c_{m,1}$. □

Localization on the selected interval.

Lemma C.18 (Single-interval localization)**.** *Let $C_{\mathrm{loc},0}$ exceed $\max\{C_{\mathrm{curv}}\kappa_{\max}^2, 2(C_0+C_1)/c_0\}$ and choose $C_{\mathrm{loc}} > C_{\mathrm{loc},0}$. Let $I=[s,e]$ be a selected interval containing exactly one change τ_j, and suppose its true split is admissible. Then every maximizer $\widehat{\tau}_j$ of $\widehat{\Gamma}_{s,e}$ satisfies*

$$d_w(\widehat{\tau}_j, \tau_j) \le C_{\mathrm{loc}} \frac{\Lambda_N}{\kappa_j^2}.$$

Proof. Consider a candidate $b < \tau_j$ and set $A=[s,b]$, $B=[b+1,\tau_j]$, and $C=[\tau_j+1,e]$. By Lemma C.1,

$$\widehat{\Gamma}_{s,e}(\tau_j) - \widehat{\Gamma}_{s,e}(b) = 2\{\widehat{D}(B,C) - \widehat{D}(A,B)\}.$$

The intervals A, B share one location, while B, C have jump κ_j. The candidate admissibility gives $S_A^w \ge m_N \ge m_{\mathrm{curv}}$, and Lemma C.17 gives $S_C^w \ge m_N$. If

$$S_B^w = d_w(b, \tau_j) \ge C_{\mathrm{loc}} \frac{\Lambda_N}{\kappa_j^2},$$

then, because $\kappa_j \le \kappa_{\max}$ and C_{loc} is large, $S_B^w \ge m_{\mathrm{curv}}$. Lemma 3.9 and (22) give

$$\widehat{D}(B,C) - \widehat{D}(A,B) \ge c_0\kappa_j^2 \frac{S_B^w S_C^w}{S_B^w + S_C^w} - C_1\Lambda_N - C_0\Lambda_N.$$

If $S_B^w \le S_C^w$, the harmonic term is at least $S_B^w/2$ and the right side is positive for sufficiently large C_{loc}. If $S_B^w > S_C^w$, Lemma C.17 gives $S_C^w \ge c_{\mathrm{cert}}\Lambda_N/\kappa_j^2$, so the harmonic term is at least $c_{\mathrm{cert}}\Lambda_N/(2\kappa_j^2)$. Choosing c_{cert} larger than a constant multiple of $(C_0+C_1)/c_0$ makes the right side positive. Therefore every candidate farther than the displayed radius has strictly smaller gain than the true split. The case $b > \tau_j$ is symmetric. □

Remark C.19. The last case in the proof is where the minimum-side certification matters. A completely jump-adaptive constant can be obtained by replacing the fixed admissibility mass by a short geometric grid of masses and maximizing over the grid with a corresponding union-bound penalty. We retain one m_N to keep the algorithm and notation readable; the stated theorem constants depend on the fixed jump range $[\kappa_{\min}, \kappa_{\max}]$.

Guard containment and preservation.

Lemma C.20 (Guard containment)**.** *If $g_N \ge C_g\Lambda_N/\kappa_{\min}^2 + w_{\max}$ with $C_g > C_{\mathrm{loc}}$, then the guard around every accepted estimate contains the corresponding true change.*

Proof. By Lemma C.18,

$$d_w(\widehat{\tau}_j, \tau_j) \le C_{\mathrm{loc}} \frac{\Lambda_N}{\kappa_j^2} \le C_{\mathrm{loc}} \frac{\Lambda_N}{\kappa_{\min}^2}.$$

The discrete guard constructed in (5) covers at least $g_N - w_{\max}$ mass on each available side of $\widehat{\tau}_j$. The choice of C_g therefore places τ_j inside the removed guard. □

Lemma C.21 (Preservation of neighboring isolation intervals)**.** *Suppose $g_N \le c_{m,2}\Delta_w$ with $c_{m,2} \le 1/12$ and the localization radius is at most $\Delta_w/12$. Removing the guard around a detected change does not remove any other true change or the weighted isolation neighborhoods of an adjacent unresolved change.*

Proof. Adjacent true changes are separated by a complete segment of mass at least Δ_w. The accepted estimate differs from its true change by at most $\Delta_w/12$. The removed guard extends by at most $g_N + w_{\max} \le \Delta_w/12 + \Delta_w/24$ beyond the estimate on either side. Thus its total reach from the detected true change is below $\Delta_w/4$. The nearest endpoint neighborhood used to isolate the next change lies within $\Delta_w/3$ of that next change and therefore at least $2\Delta_w/3$ from the detected change before discretization. The guard and that neighborhood are disjoint. The same reasoning applies on the left. □

Proof of Theorem 3.11.

Proof. Let $\mathcal{E}$ be the intersection of the uniform score event, empirical-curvature event, empirical merge-cost event, and weighted isolation event. The preceding lemmas and their union bounds give

$$\mathbb{P}(\mathcal{E}) \geq 1 - \delta$$

after distributing the failure budget and adjusting numerical constants.

We argue by induction over recursive calls on $\mathcal{E}$. Initially, every true change is unresolved and has an isolating interval. If a search range contains no unresolved change, every retained interval is homogeneous after previously detected guards have been removed, so Lemma C.14 makes the over-threshold set empty and recursion stops.

If the range contains at least one unresolved change, Lemma C.21 ensures that an isolating interval for such a change remains inside the range. Lemma C.15 makes the over-threshold set nonempty. Lemma C.16 shows that the selected interval contains exactly one unresolved change, and Lemma C.17 makes its true split admissible. Lemma C.18 yields (17) for the accepted estimate.

Lemma C.20 removes the corresponding true change from both child search ranges, so it cannot be detected twice. Lemma C.21 shows that no other change or isolating interval is removed. Therefore the two child calls partition the remaining unresolved changes without loss. Each accepted split reduces their number by one. After exactly K acceptances, all child ranges contain no unresolved change and stop by the no-false-positive lemma. Hence $\widehat{K} = K$, and every accepted boundary satisfies the stated localization bound. □

C.3 Extensions, Lower Bounds, and Model Verification

This appendix develops results that complement the main recovery theorem: stability to estimated weights, an explicitly delimited information lower bound, reduction identities, and concrete conditions under which the curvature assumptions hold for common score distributions and contamination models. Let w_i° be the oracle capped weights in (2) and define

$$\mathcal{E}_w = \{c_w w_i^\circ \leq \widehat{w}_i \leq C_w w_i^\circ \text{ for all } i \in [N]\},$$

where $0 < c_w \leq C_w < \infty$.

Corollary 3.12 is stated in Section 3; we now prove it.

Proof of Corollary 3.12. Condition on the sigma-field that generates $\widehat{w}_{1:N}$ and on $\mathcal{E}_w$. For every interval A,

$$c_w S_A^{w^\circ} \leq S_A^{\widehat{w}} \leq C_w S_A^{w^\circ}.$$

Similarly, for every pair u, v,

$$c_w d_{w^\circ}(u, v) \leq d_{\widehat{w}}(u, v) \leq C_w d_{w^\circ}(u, v). \tag{34}$$

Thus the estimated-weight spacing is at least $c_w \Delta_{w^\circ}$, and the no-dominant-atom ratio changes by at most C_w/c_w. The weighted-coordinate measure of every isolation neighborhood changes by the same constant factors, so the interval count in Lemma 3.10 is multiplied by at most a constant depending on (c_w, C_w).

By the conditional assumptions, population Huber risks computed with $\widehat{w}_i$ satisfy the same centering and curvature inequalities with constants bounded away from zero and infinity uniformly on $\mathcal{E}_w$. Theorem 3.11 therefore applies conditionally and gives

$$d_{\widehat{w}}(\widehat{\tau}_j, \tau_j) \leq C \frac{\Lambda_N}{\kappa_j^2}$$

with conditional probability at least $1 - \delta$. The first inequality in (34) implies

$$d_{w^\circ}(\widehat{\tau}_j, \tau_j) \leq c_w^{-1} d_{\widehat{w}}(\widehat{\tau}_j, \tau_j).$$

Finally,

$$\mathbb{P}(\text{failure}) \leq \mathbb{P}(\mathcal{E}_w^c) + \mathbb{P}(\text{failure} \mid \mathcal{E}_w) \leq \delta_w + \delta.$$

□

A sample-splitting construction. The comparability event can be verified in simple calibration designs. Suppose an auxiliary corpus supplies R independent score replicates for each length bin g. Let v_g be the bin variance proxy and define a robust scale estimate $\widehat{v}_g$ from the auxiliary data. If

$$\mathbb{P}\left(\max_{g\in[G]}\left|\frac{\widehat{v}_g}{v_g}-1\right|>\eta\right)\le\delta_w,\qquad 0<\eta<1,$$

then, before capping,

$$\frac{1}{1+\eta}v_g^{-1}\le\widehat{v}_g^{-1}\le\frac{1}{1-\eta}v_g^{-1}. \tag{35}$$

Projection onto a common interval $[w_{\min},w_{\max}]$ is monotone and nonexpansive, so capped weights satisfy a constant-factor version of (35). Because the auxiliary corpus is independent of the target sequence, conditioning is immediate.

C.3.1 Information lower bound

To calibrate this upper rate, consider the independent one-change Gaussian submodel $Y_i\sim N(0,\sigma_i^2)$ before the boundary and $Y_i\sim N(\kappa,\sigma_i^2)$ after it, with information distance $d_{\mathrm{info}}(u,v)=\sum_{i\in(u,v]\cup(v,u]}\sigma_i^{-2}$. Let $Q_{\mathrm{info}}(\delta,\widehat{\tau},P_\tau)$ be the smallest radius containing $\widehat{\tau}$ with probability at least $1-\delta$.

Theorem C.22 (Two-point information lower bound)**.** *Fix $\delta\in(0,1/8)$ and $c_H\in(0,1/2]$. If the admissible class contains $\tau_0<\tau_1$ with $H=d_{\mathrm{info}}(\tau_0,\tau_1)$ satisfying*

$$c_H\frac{\log(1/\delta)}{\kappa^2}\le H\le\frac{1}{2}\frac{\log(1/\delta)}{\kappa^2}, \tag{36}$$

then

$$\inf_{\widehat{\tau}}\sup_{\tau}Q_{\mathrm{info}}(\delta,\widehat{\tau},P_\tau)\ge\frac{c_H}{3}\frac{\log(1/\delta)}{\kappa^2}.$$

Corollary C.23 (Independent-regime rate comparison)**.** *If caps are inactive, $w_i\asymp\sigma_i^{-2}$, L_N is polynomial in N, and $\delta=N^{-a}$, the RWCP upper bound and Theorem C.22 are both of order $\log N/\kappa^2$, up to constants. This comparison is restricted to the admissible independent Gaussian submodel and is not asserted under dependence or active caps.*

For a fixed variance sequence $\sigma_{1:N}^2$ and an admissible set of locations $\mathcal{T}_N$, let P_τ denote the independent Gaussian model

$$Y_i\sim\begin{cases}N(0,\sigma_i^2), & i\le\tau,\\ N(\kappa,\sigma_i^2), & i>\tau.\end{cases}$$

For an estimator $\widehat{\tau}$, define its $(1-\delta)$ information-radius by

$$Q_{\mathrm{info}}(\delta,\widehat{\tau},P_\tau)=\inf\left\{r\ge0:P_\tau(d_{\mathrm{info}}(\widehat{\tau},\tau)\le r)\ge1-\delta\right\}.$$

Theorem C.22 and Corollary C.23 are stated in Section 3; the following testing reduction proves them.

A testing reduction.

Lemma C.24 (Separated localization implies testing)**.** *Let P_0,P_1 have change points τ_0,τ_1 separated by*

$$H=d_{\mathrm{info}}(\tau_0,\tau_1).$$

If an estimator satisfies

$$P_j\left(d_{\mathrm{info}}(\widehat{\tau},\tau_j)<\frac{H}{3}\right)\ge1-\delta,\qquad j=0,1,$$

then there exists a test between P_0 and P_1 whose sum of type-I and type-II errors is at most 2δ.

Proof. The information distance is a path metric on ordered indices. Therefore the open balls

$$B_j=\{t:d_{\mathrm{info}}(t,\tau_j)<H/3\}$$

are disjoint: if t belonged to both, the triangle inequality would give $H<2H/3$. Define the test $\varphi=1$ when $\widehat{\tau}\in B_1$ and $\varphi=0$ otherwise. Under P_0, a type-I error implies $\widehat{\tau}\notin B_0$, so its probability is at most δ. Under P_1, a type-II error implies $\widehat{\tau}\notin B_1$, also with probability at most δ. □

Lemma C.25 (High-probability two-point bound)**.** *For any two distributions P_0, P_1 and any test φ,*

$$P_0(\varphi = 1) + P_1(\varphi = 0) \geq \frac{1}{2} \exp\{-D_{\mathrm{KL}}(P_0 \| P_1)\}.$$

Proof. Let $L = \mathrm{d}P_0/\mathrm{d}P_1$ on the absolutely continuous part. The sum of testing errors is at least

$$\int \min(\mathrm{d}P_0, \mathrm{d}P_1).$$

The Bretagnolle–Huber inequality states

$$\int \min(\mathrm{d}P_0, \mathrm{d}P_1) \geq \frac{1}{2} e^{-D_{\mathrm{KL}}(P_0 \| P_1)}.$$

For completeness, write $A = \{L \geq 1\}$. Jensen's inequality under P_0 on A^c and under P_1 on A bounds the overlap from below by the displayed exponential quantity; equivalently the result follows from the variational representation of KL divergence applied to the binary partition (A, A^c). □

Proof of Theorem C.22.

Proof. Let $P_0 = P_{\tau_0}$ and $P_1 = P_{\tau_1}$ for the two admissible locations in (36). The two distributions differ only on $(\tau_0, \tau_1]$, and Gaussian additivity gives

$$D_{\mathrm{KL}}(P_0 \| P_1) = \frac{\kappa^2}{2} \sum_{i=\tau_0+1}^{\tau_1} \sigma_i^{-2} = \frac{\kappa^2 H}{2} \leq \frac{1}{4} \log(1/\delta). \tag{37}$$

If one estimator had information-radius strictly below $H/3$ under both models, Lemma C.24 would yield a test with total error at most 2δ. Lemma C.25 and (37) instead give total error at least $\frac{1}{2}\delta^{1/4} > 2\delta$ for $\delta < 1/8$, a contradiction. Hence one of the two models has radius at least $H/3$, and the lower bound in (36) completes the proof. □

Index and token interpretations.

Corollary C.26 (Index-distance lower bound)**.** *If $0 < \underline{\sigma}^2 \leq \sigma_i^2 \leq \overline{\sigma}^2 < \infty$, then*

$$|u - v|/\overline{\sigma}^2 \leq d_{\mathrm{info}}(u, v) \leq |u - v|/\underline{\sigma}^2.$$

Consequently Theorem C.22 implies an index localization lower bound of order $\underline{\sigma}^2 \log(1/\delta)/\kappa^2$.

Proof. The path between u and v contains $|u - v|$ atoms, each between $1/\overline{\sigma}^2$ and $1/\underline{\sigma}^2$. Sum these inequalities and apply Theorem C.22. □

Corollary C.27 (Token-distance interpretation)**.** *If $c_1 n_i^{-1} \leq \sigma_i^2 \leq c_2 n_i^{-1}$, then*

$$\frac{1}{c_2} \sum_{i \in (u,v] \cup (v,u]} n_i \leq d_{\mathrm{info}}(u, v) \leq \frac{1}{c_1} \sum_{i \in (u,v] \cup (v,u]} n_i.$$

Thus the minimax information lower bound is equivalent, up to constants, to a lower bound on the number of misplaced tokens.

C.3.2 Reduction results

Proof of Proposition 2.2. Write $x = S^w_{s:b}$, $y = S^w_{b+1:e}$, and let $\bar{Y}_L, \bar{Y}_R, \bar{Y}_P$ be the weighted means of the left, right, and parent intervals. Completing the square gives the unconstrained gain $xy(\bar{Y}_L - \bar{Y}_R)^2/(x + y)$. If the fit is constrained to Θ, the gain instead equals

$$\{W^Y_{s,e}(b)\}^2 + (x + y)d(\bar{Y}_P, \Theta)^2 - xd(\bar{Y}_L, \Theta)^2 - yd(\bar{Y}_R, \Theta)^2.$$

Thus the stated identity holds whenever the projection terms vanish. For example, $Y = (2, 3)$, unit weights, and $\Theta = [-1, 1]$ give constrained gain zero but squared weighted CUSUM $1/2$. □

Proposition C.28 (Equal-weight reduction)**.** *If $w_i \equiv w$, weighted-coordinate sampling reduces to index-uniform sampling up to endpoint discretization, $d_w(u,v) = w|u-v|$, and RWCP becomes a robust profile-loss version of unweighted change point detection.*

Proposition C.29 (Additive-score WCP–GCP equivalence)**.** *Suppose a segment-level detector obeys*

$$\phi(X_{a:b}) = \frac{\sum_{i=a}^{b} \nu_i \phi(X_i)}{\sum_{i=a}^{b} \nu_i} \tag{38}$$

for known $\nu_i > 0$. Then the generalized CUSUM based on $\phi(X_{s:t})$ and $\phi(X_{t+1:e})$ equals WCP with weights $w_i = \nu_i$. Without (38)*, exact equivalence need not hold, particularly for context-sensitive detectors applied to concatenated text.*

Proof of Proposition C.28. If $w_i \equiv w$, then $W(t) = wt$ and weighted-coordinate uniform sampling maps to uniform sampling over the index axis up to the discretization created by W^{-1}. Cumulative masses satisfy $S^w_{a:b} = w(b-a+1)$ and $d_w(u,v) = w|u-v|$. The Huber loss becomes

$$\sum_{i \in A} \rho_c\{\sqrt{w}(Y_i - \theta)\},$$

which differs from an unweighted robust profile loss only by the common scale $\sqrt{w}$ and the corresponding tuning convention. Hence the method is a robust profile-loss analogue of VCP. □

Proof of Proposition C.29. Under (38),

$$\phi(X_{s:b}) = \bar{Y}^{\nu}_{s:b}, \qquad \phi(X_{b+1:e}) = \bar{Y}^{\nu}_{b+1:e}.$$

Substitution into the generalized segment contrast gives

$$\left(\frac{S^{\nu}_{s:b} S^{\nu}_{b+1:e}}{S^{\nu}_{s:e}}\right)^{1/2} |\bar{Y}^{\nu}_{s:b} - \bar{Y}^{\nu}_{b+1:e}|,$$

which is precisely WCP with $w_i = \nu_i$. If the detector evaluates concatenated text with cross-sentence context, (38) may fail, and no algebraic identity follows. □

C.3.3 Assumption verification

Proposition C.30 (Curvature from central density)**.** *Suppose $Y_i = \theta_j + \sigma_i \xi_i$, $a_i = \sqrt{w_i}\sigma_i \in [a_-, a_+]$, $\mathbb{E}\rho_c(a_i\xi_i) < \infty$, $\mathbb{E}\psi_c(a_i\xi_i) = 0$ (for example by symmetry), and ξ_i has density f_i satisfying*

$$f_i(x) \geq f_0 > 0 \quad \text{for } |x| \leq R.$$

If $c/a_+ + r_0/(\sigma_i) \leq R$ uniformly, then Assumption 3.2 holds with

$$m_0 \geq 2 f_0 \frac{c}{a_+}$$

up to truncation at one, and $M_0 \leq 1$.

Proof. At differentiability points,

$$\begin{aligned} Q_i''(\theta) &= w_i \mathbb{P}\left(|\sqrt{w_i}(Y_i - \theta)| < c\right) \\ &= w_i \mathbb{P}\left(|a_i\xi_i + \sqrt{w_i}(\theta_j - \theta)| < c\right). \end{aligned}$$

For $|\theta - \theta_j| \leq r_0$, the event contains an interval in ξ_i of length at least $2c/a_+$ lying inside $[-R, R]$. Its probability is at least $2f_0 c/a_+$. The upper bound follows because a probability is at most one. The score-centering condition makes θ_j a minimizer; only then does integrating the second-derivative bounds twice around θ_j give (7). □

Proposition C.31 (Active caps change the metric)**.** *Let $w_i^{\circ} = \Pi_{[w_{\min}, w_{\max}]}(\sigma_i^{-2})$. Then*

$$\sum_{i \in (u,v]} w_i^{\circ} \leq d_{\mathrm{info}}(u,v) \tag{39}$$

need not hold if the lower cap is active, and the reverse inequality need not hold if the upper cap is active. Therefore a guarantee in $d_{w^{\circ}}$ cannot generally be called inverse-variance minimax optimal.

Proof. If $\sigma_i^{-2} < w_{\min}$, then $w_i^{\circ} > w_i^{\mathrm{info}}$, violating (39). If $\sigma_i^{-2} > w_{\max}$, then $w_i^{\circ} < w_i^{\mathrm{info}}$, violating the reverse comparison. Only uniform comparability or inactive caps identify the two metrics up to constants. □

C.3.4 DISTRIBUTIONAL AND CONTAMINATION CALCULATIONS

This section makes the abstract curvature constants in the main theorem concrete. The calculations are not needed for the logical validity of Theorem 3.11, but they clarify which score distributions satisfy its assumptions and how the Huber parameter affects the constants. Throughout this section, let

$$q_F(u) = \mathbb{E}_F \rho_c(Z-u), \qquad Z \sim F,$$

where F is a standardized detector-noise distribution. Define the central-mass modulus

$$m_F(c,r) = \inf_{|u|\le r} \mathbb{P}_F(|Z-u| < c).$$

The next proposition converts this probability into both a risk-curvature bound and a two-population separation bound.

Proposition C.32 (Exact standardized Huber curvature)**.** *Suppose F is symmetric around zero and $\mathbb{E}_F|Z| < \infty$. Then $u = 0$ minimizes q_F. Moreover, at every differentiability point,*

$$q_F'(u) = -\mathbb{E}_F \psi_c(Z-u), \qquad q_F''(u) = \mathbb{P}_F(|Z-u| < c).$$

Consequently, for every $|u| \le r$,

$$\frac{m_F(c,r)}{2}u^2 \le q_F(u) - q_F(0) \le \frac{1}{2}u^2. \tag{40}$$

Let $B, C > 0$ and $|\kappa| \le r$. Define the two-population profile separation

$$\mathfrak{D}_F(B,C;\kappa) = 2\inf_{t\in\mathbb{R}} \left[B\{q_F(t) - q_F(0)\} + C\{q_F(t-\kappa) - q_F(0)\}\right]. \tag{41}$$

Then

$$m_F(c,r)\frac{BC}{B+C}\kappa^2 \le \mathfrak{D}_F(B,C;\kappa) \le \frac{BC}{B+C}\kappa^2. \tag{42}$$

Proof. The Huber loss is convex and differentiable except at two points. Dominated differentiation is valid for the first derivative because $|\psi_c| \le c$. Symmetry of F and oddness of ψ_c imply

$$q_F'(0) = -\mathbb{E}_F \psi_c(Z) = 0.$$

Convexity therefore makes zero a global minimizer. At points where F has no atom at $u \pm c$, differentiating once more gives

$$q_F''(u) = \mathbb{E}_F \mathbf{1}\{|Z-u| < c\} = \mathbb{P}_F(|Z-u| < c).$$

The same identity holds in the distributional sense when boundary atoms are present. For $u \ge 0$, the fundamental theorem of calculus and $q_F'(0) = 0$ yield

$$q_F(u) - q_F(0) = \int_0^u q_F'(v)\,\mathrm{d}v = \int_0^u \int_0^v q_F''(z)\,\mathrm{d}z\,\mathrm{d}v.$$

On $[0,r]$, the inner integrand lies between $m_F(c,r)$ and one. This proves (40); the case $u < 0$ follows by the same argument or by symmetry.

Because q_F is convex with unique minimizer zero whenever $m_F(c,r) > 0$, a minimizer of the expression in (41) lies between 0 and κ when $\kappa > 0$, and between κ and 0 when $\kappa < 0$. Hence both t and $t-\kappa$ have absolute value at most r. Applying the lower bound in (40) gives

$$\begin{aligned}\mathfrak{D}_F(B,C;\kappa) &\ge m_F(c,r)\inf_t\{Bt^2 + C(t-\kappa)^2\} \\ &= m_F(c,r)\frac{BC}{B+C}\kappa^2,\end{aligned}$$

where the minimizing quadratic argument is $t = C\kappa/(B+C)$. Evaluating the original objective at this same argument and using the upper bound in (40) proves the upper inequality. □

Remark C.33 (Why the profile formulation is useful). Proposition C.32 controls the loss paid by forcing two populations to share one location. It does not require an explicit formula for the minimizer of the mixed risk

$$t \mapsto Bq_F(t) + Cq_F(t-\kappa).$$

For a nonquadratic Huber loss, that minimizer is generally not $C\kappa/(B+C)$. The quadratic point is used only as a feasible argument for the upper bound, while strong convexity supplies the lower bound. This is precisely the distinction that makes profile-risk geometry more tractable than a direct difference of mixed-segment Huber centers.

Gaussian and Student examples. The central-mass modulus has a closed form for several standard heavy- and light-tailed reference families.

Corollary C.34 (Gaussian curvature constant)**.** *If $Z \sim N(0,1)$ and Φ denotes the standard normal distribution function, then*

$$m_{\mathrm{G}}(c,r) = \Phi(c-r) + \Phi(c+r) - 1.$$

Therefore, for every $|\kappa| \le r$,

$$\{\Phi(c-r) + \Phi(c+r) - 1\}\,\frac{BC}{B+C}\kappa^2 \le \mathfrak{D}_{\mathrm{G}}(B,C;\kappa) \le \frac{BC}{B+C}\kappa^2.$$

Proof. For $u \ge 0$,

$$\begin{aligned}\mathbb{P}(|Z-u| < c) &= \Phi(c+u) - \Phi(-c+u) \\ &= \Phi(c+u) + \Phi(c-u) - 1. \end{aligned} \tag{43}$$

Its derivative is $\varphi(c+u) - \varphi(c-u) \le 0$, because $|c-u| \le c+u$ and the Gaussian density φ decreases with absolute value. The probability is even in u, so its minimum on $[-r,r]$ is attained at $|u| = r$. Substitute the resulting modulus into Proposition C.32. □

Corollary C.35 (Student curvature constant)**.** *Let Z have a centered Student distribution with $\nu > 1$ degrees of freedom and distribution function T_ν. Then*

$$m_{t_\nu}(c,r) = T_\nu(c-r) + T_\nu(c+r) - 1, \tag{44}$$

and the separation sandwich (42) *holds with this constant.*

Proof. The Student density is symmetric and nonincreasing in absolute value. Repeating (43) with T_ν in place of Φ shows that the probability of a length-$2c$ window is minimized at the largest displacement $|u| = r$. Since $\nu > 1$, the first moment required in Proposition C.32 is finite. □

The preceding corollary illustrates why the raw detector score need not be sub-Gaussian. A Student variable with small ν has polynomial tails, while the Huber score remains bounded and the local population risk still has positive curvature whenever (44) is positive.

Symmetric and asymmetric contamination. We next quantify two distinct effects of contamination. Symmetric contamination preserves the Huber target but reduces curvature. Asymmetric contamination can move the target, although bounded Huber influence controls that displacement.

Proposition C.36 (Symmetric contamination preserves the target)**.** *Let*

$$F_\epsilon = (1-\epsilon)F_0 + \epsilon H, \qquad 0 \le \epsilon < 1, \tag{45}$$

where both F_0 and H are symmetric around zero and have finite first moments. Then zero minimizes q_{F_ϵ} and

$$m_{F_\epsilon}(c,r) \ge (1-\epsilon)m_{F_0}(c,r). \tag{46}$$

Consequently,

$$\mathfrak{D}_{F_\epsilon}(B,C;\kappa) \ge (1-\epsilon)m_{F_0}(c,r)\frac{BC}{B+C}\kappa^2$$

for every $|\kappa| \le r$.

Proof. A mixture of symmetric distributions is symmetric, so the first conclusion follows from Proposition C.32. For every u,

$$\begin{aligned}\mathbb{P}_{F_\epsilon}(|Z-u| < c) &= (1-\epsilon)\mathbb{P}_{F_0}(|Z-u| < c) + \epsilon\mathbb{P}_H(|Z-u| < c) \\ &\ge (1-\epsilon)\mathbb{P}_{F_0}(|Z-u| < c).\end{aligned}$$

Taking the infimum over $|u| \le r$ proves (46); the separation result follows from (42). □

Proposition C.37 (Displacement under asymmetric contamination)**.** *Let F_0 be symmetric around zero and suppose*

$$q_{F_0}(u) - q_{F_0}(0) \geq \frac{m_0}{2}u^2, \qquad |u| \leq r,$$

for $m_0 > 0$. Let H be an arbitrary distribution with finite first moment and define F_ϵ by (45)*. Let*

$$u_\epsilon \in \arg\min_{|u|\leq r} q_{F_\epsilon}(u).$$

Then

$$|u_\epsilon| \leq \frac{2\epsilon c}{(1-\epsilon)m_0}.$$

If the right-hand side is strictly smaller than r, every constrained minimizer is interior and hence is also an unconstrained local minimizer.

Proof. The Huber loss is c-Lipschitz, so for every u,

$$|q_H(u) - q_H(0)| \leq c|u|.$$

Because u_ϵ minimizes the contaminated risk over $[-r, r]$,

$$\begin{aligned} 0 &\geq q_{F_\epsilon}(u_\epsilon) - q_{F_\epsilon}(0) \\ &= (1-\epsilon)\{q_{F_0}(u_\epsilon) - q_{F_0}(0)\} + \epsilon\{q_H(u_\epsilon) - q_H(0)\} \\ &\geq \frac{(1-\epsilon)m_0}{2}u_\epsilon^2 - \epsilon c|u_\epsilon|. \end{aligned}$$

If $u_\epsilon = 0$, the conclusion is immediate. Otherwise divide by $|u_\epsilon|$ and rearrange. Strict interiority follows when the bound is less than r. □

Remark C.38 (Interpretation of asymmetric contamination). Theorem 3.11 is formulated in terms of the segment-wise population Huber locations actually identified by the score distributions. Proposition C.37 therefore does not add an unmodeled bias term to that theorem. Instead, it quantifies how far those robust targets may move from a symmetric clean center. If two adjacent segments have clean centers separated by κ_0 and contamination levels ϵ_- and ϵ_+, then their contaminated Huber-location jump obeys the deterministic lower bound

$$\kappa \geq \left[\kappa_0 - \frac{2\epsilon_- c}{(1-\epsilon_-)m_-} - \frac{2\epsilon_+ c}{(1-\epsilon_+)m_+}\right]_+.$$

Thus contamination affects localization through both curvature and effective separation.

Finite-sample replacement sensitivity. The next result is deterministic. It does not claim robustness to an arbitrary fraction of adversarially replaced sentences; rather, it quantifies the local influence of a finite replacement set on one interval-level Huber fit.

Proposition C.39 (Deterministic replacement sensitivity)**.** *Fix an interval A and observations $\{y_i : i \in A\}$. Let*

$$L_A(\theta) = \sum_{i\in A} \rho_c(\sqrt{w_i}(y_i - \theta))$$

have minimizer $\widehat{\theta}_A$. Replace the observations on a set $C \subseteq A$ by arbitrary values to obtain $\widetilde{L}_A$ and minimizer $\widetilde{\theta}_A$. Suppose both minimizers lie in an interval I on which L_A is αS_A^w-strongly convex. Then

$$|\widetilde{\theta}_A - \widehat{\theta}_A| \leq \frac{2c}{\alpha S_A^w}\sum_{i\in C}\sqrt{w_i} \leq \frac{2c\sqrt{w_{\max}|C|}}{\alpha S_A^w}. \tag{47}$$

Proof. Choose zero subgradients $g_A(\widehat{\theta}_A) = 0$ and $\widetilde{g}_A(\widetilde{\theta}_A) = 0$. Replacing one observation changes its score contribution

$$\sqrt{w_i}\psi_c(\sqrt{w_i}(y_i - \theta))$$

by at most $2c\sqrt{w_i}$, uniformly in θ. Hence

$$|g_A(\theta) - \widetilde{g}_A(\theta)| \leq 2c\sum_{i\in C}\sqrt{w_i} \qquad \text{for all } \theta.$$

Strong convexity implies strong monotonicity of the subgradient map on I:

$$|g_A(u) - g_A(v)| \geq \alpha S_A^w |u - v|, \qquad u, v \in I.$$

Apply this inequality with $u = \widetilde{\theta}_A$ and $v = \widehat{\theta}_A$. Since $g_A(\widehat{\theta}_A) = 0$ and $\widetilde{g}_A(\widetilde{\theta}_A) = 0$,

$$\begin{aligned} \alpha S_A^w |\widetilde{\theta}_A - \widehat{\theta}_A| &\leq |g_A(\widetilde{\theta}_A)| \\ &= |g_A(\widetilde{\theta}_A) - \widetilde{g}_A(\widetilde{\theta}_A)| \\ &\leq 2c \sum_{i \in C} \sqrt{w_i}. \end{aligned}$$

The second inequality in (47) follows from $w_i \leq w_{\max}$. □

Remark C.40 (Scope of Proposition C.39). The bound concerns the fitted location on a fixed interval and requires empirical curvature along the path between the two fits. It does not by itself establish exact change point recovery under adversarial replacements, because an adversary may also alter which intervals cross the threshold. A full adversarial theorem would need a simultaneous bound on every affected profile gain and a contamination-aware signal condition. We therefore use (47) as a sensitivity diagnostic rather than as an unsupported global robustness claim.

Constants and the Huber tuning parameter. The main theorem suppresses fixed constants to keep its statement readable. Tracking the proof reveals the two roles of c. The score concentration scale is proportional to the uniform score bound

$$b_c = c\sqrt{w_{\max}},$$

whereas empirical and population curvature are controlled by central probabilities such as $m_F(c, r)$ and $p_0(c, r)$. The resulting localization constant therefore has the schematic form

$$d_w(\widehat{\tau}_j, \tau_j) \leq C_{\mathrm{dep},w}\, \mathfrak{K}(c, r, p_0) \frac{\Lambda_N(\delta)}{\kappa_j^2}, \tag{48}$$

where $C_{\mathrm{dep},w}$ depends on fixed dependence and weight bounds, and $\mathfrak{K}(c, r, p_0)$ collects the bounded-score and inverse-curvature constants. We do not claim this schematic factor is the sharp optimized constant. Crucially, all multiplicity and dependence growth remains in $\Lambda_N(\delta)$, including the q_N^2 penalty. Equation (48) records the roles of the tuning quantities without replacing the rate in Theorem 3.11.

For a reference family F, a transparent tuning proxy is

$$\mathfrak{C}_F(c; r) = \frac{c^2}{m_F(c, r)^2}.$$

A small value of c limits the influence of extreme scores but may reduce the mass in the locally quadratic region. A large value increases curvature toward its quadratic-loss limit but also increases the bounded-score constant. The proxy can be minimized on a predetermined compact grid without using ground-truth boundaries. In practice, the grid should be chosen using a training or calibration corpus so that selection does not invalidate the target-document guarantee.

Proposition C.41 (Continuity of the tuning proxy)**.** *Suppose F has a continuous strictly positive density on every compact interval. For fixed $r < \infty$, the function $c \mapsto m_F(c, r)$ is continuous and strictly positive for $c > 0$. Hence $\mathfrak{C}_F(c; r)$ is continuous on every compact interval $[c_-, c_+] \subset (0, \infty)$ and attains a minimum there.*

Proof. For each fixed u, the map

$$c \mapsto \mathbb{P}_F(|Z - u| < c) = F(u + c) - F(u - c)$$

is continuous. Joint continuity in (c, u) follows from continuity of the distribution function, and the infimum over the compact set $|u| \leq r$ is therefore continuous by the maximum theorem. Strict positivity follows because every interval $(u - c, u + c)$ has positive probability under a strictly positive density. The remaining claims follow from elementary continuity and compactness. □

Remark C.42 (Quadratic-loss limit). The algebraic identity in Proposition 2.2 holds in the quadratic limit when the unconstrained means are feasible. The heavy-tail concentration proof, however, does not pass to this limit because the score bound b_c diverges. Recovering the classical WCP probability bound at $c = \infty$ requires a separate tail assumption on the raw scores, such as the sub-Gaussian condition used in the foundational analysis. Algebraic reduction and probabilistic reduction are therefore distinct statements.