Online Supplement: Tipping Fractions for Regression Conclusions

Lucas D. Konrad and Nikolas Kuschnig

November 24, 2026

We begin with the finite-type and independent-stratified tipping inference used for Mindspark. The next two sections give inference for the sensitivity curve and its proofs, including the connection to tipping inference. The remaining sections give sufficient conditions and population examples, calibration designs and evidence, and additional sample analyses for Mindspark and STAR. The main paper's appendix contains the scalar tipping proofs, the curve characterization and computation, empirical specifications, primary scalar calibration methods and finite-sample limitations. References to results outside this supplement refer to the main paper.

S1 Tipping Interval for Mindspark

With the released Mindspark scoring map held fixed, the observed-case population has finitely many record types (Section D.1). We maintain that tipping splits the mass of a single type at the cutoff, rather than approximating the scores by a continuous distribution. Corollary S1 shows that under independent student records and fixed lottery counts within strata, the iid variance is asymptotically conservative for the observed-case mixture.

Proposition S1 (Singleton cutoff)**.** *Let P assign positive probabilities λ_j to finitely many distinct records $o_j = (y_j, w_j, z_j)$. Suppose Q_z is nonsingular, $L > 0$, $c = \vartheta - b > 0$ and $P(q_b < 0) > 0$, with ordinary least-squares adjustments and either a fixed floor or $b = \kappa\vartheta$, $0 \le \kappa < 1$. Write $q_j = q_b(o_j)$ and $m = \sum_j \lambda_j q_j$. Assume some type j_* has score*

$a = q_{j_*} > 0$ *different from every other type's score and, with* $H = \{j : q_j > a\}$,

$$\sum_{j \in H} \lambda_j q_j < m < \sum_{j \in H} \lambda_j q_j + \lambda_{j_*} a.$$

Thus the crossing deletes all of H *and mass* $\rho = p^\dagger - \sum_{j \in H} \lambda_j$ *strictly between* 0 *and* λ_{j_*} *from the cutoff type. Let* D *be the fractional deletion rule:* $D = 1$ *on* H, $D = \rho/\lambda_{j_*}$ *on* j_* *and* $D = 0$ *elsewhere. Use it in the loadings of Section A, and suppose the resulting complete influence* φ_A *in* (5) *has positive variance. Under iid sampling,*

$$\sqrt{N}(\hat{p}^\dagger - p^\dagger) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \psi(O_i) + o_p(1), \qquad \psi = -\varphi_A / a.$$

The fitted complete variance is consistent, the integer share $k^\dagger/N$ *has the same limit, and the grid-cell interval, before any refusal checks, covers the clipped crossing asymptotically with probability at least* $1 - \alpha$.

Proof. All fitted population ingredients are smooth functions of the finite vector λ near the reference law. The strict inequalities preserve H, the singleton cutoff and the interior split locally, giving the exact formula

$$p^\dagger(\lambda') = \sum_{j \in H} \lambda'_j + \frac{m(\lambda') - \sum_{j \in H} \lambda'_j q_j(\lambda')}{q_{j_*}(\lambda')}.$$

At fixed p near $p^\dagger$, the balance is likewise smooth in λ' and affine in p. Differentiating it gives the centered known-score term $\ell_a(q_b) - \mathbb{E}[\ell_a(q_b)]$ (with $\ell_a(q) = (q-a)_+ - q$) and the adjustment term $\mathbb{E}[(D-1)\partial_\nu q_b]'\phi_\nu$, where $\nu = (\pi, \eta, b)$ and ϕ_ν is the joint least-squares influence. Projection orthogonality reduces the adjustment term to the loadings in Section A; the slope in p is the marginal support a. The multinomial delta method proves the expansion, and $0 \le k^\dagger/N - \hat{p}^\dagger < 1/N$ transfers the limit. The fitted cutoff remains the same type and its split fraction converges, so the fitted loadings and complete variance are consistent. For a change $\Delta = \lambda' - \lambda$, the balance expansion using the influence at $p^\dagger$ has remainder $O(\|\Delta\|^2 + |p - p^\dagger|\,\|\Delta\|)$ by smoothness of the frequency derivative in p. Hence the expansion and fitted variance consistency hold uniformly on shrinking neighborhoods of the crossing. The endpoint argument of Corollary 1 gives coverage; for a crossing outside the reporting range, consistency and the strictly signed boundary balance suffice. □

Fixed-count stratification. Let M students have independent potential records, identically distributed within each of finitely many strata s. Assignment is independent of those records with fixed stratum-by-offer counts $M_{sa} = Mr_{sa} + O(1)$, where $r_{sa} > 0$ and $\sum_{s,a} r_{sa} = 1$. Let $R(a)$ indicate usable assessments under offer a, allowing response to depend on the student's potential record; potential records include these response indicators and are independent across students. For $\omega = \sum_{s,a} r_{sa}\mathbb{E}_s[R(a)] > 0$, define the observed-case target law by

$$P_{\text{obs}}(B) = \frac{1}{\omega}\sum_{s,a} r_{sa}\mathbb{E}_s[R(a)\mathbf{1}\{(Y(a), a, Z) \in B\}]\,.$$

Corollary S1 (Conservative iid variance)**.** *Suppose P_{obs} satisfies Proposition S1, and let ψ be its crossing influence. With $N_{\text{obs}} = \sum_i R_i$ and $\mu_{sa} = \mathbb{E}_s[R(a)\psi(Y(a), a, Z)]$,*

$$\sqrt{N_{\text{obs}}}(\hat{p}^\dagger - p^\dagger) \xrightarrow{d} \mathrm{N}(0, V_{\text{strat}}), \qquad V_{\text{strat}} = \frac{1}{\omega}\sum_{s,a} r_{sa}\mathrm{Var}_s\{R(a)\psi\}.$$

The iid variance satisfies

$$V_{\text{iid}} = \mathbb{E}_{P_{\text{obs}}}[\psi^2] = V_{\text{strat}} + \frac{1}{\omega}\sum_{s,a} r_{sa}\mu_{sa}^2 \geq V_{\text{strat}}.$$

Before any refusal checks, the same grid-cell interval using the iid complete-influence standard errors has asymptotic coverage at least $1 - \alpha$ for this law's clipped crossing.

Proof. Empirical type proportions converge to those of P_{obs} at rate $O_p(M^{-1/2})$, so the same smooth finite-dimensional map gives $\hat{p}^\dagger - p^\dagger = N_{\text{obs}}^{-1}\sum_i R_i\psi(O_i) + o_p(M^{-1/2})$. Conditional on assignment, the summands are independent and bounded, their total mean is $\sum_{s,a} M_{sa}\mu_{sa} = O(1)$ since $\sum_{s,a} r_{sa}\mu_{sa} = 0$, and their variance is $M\sum_{s,a} r_{sa}\mathrm{Var}_s(R(a)\psi) + O(1)$. The central limit theorem and $N_{\text{obs}}/M \to \omega$ give the stated limit; expanding the cell second moments gives the variance identity. Since the padded grid has mesh $1/N_{\text{obs}} = o_p(1)$, its tested endpoints in the cell containing a crossing in the reporting range approach $p^\dagger$ in probability. The uniform expansion and variance consistency therefore give the centered, iid-studentized limit at the realized endpoints, with variance $V_{\text{strat}}/V_{\text{iid}} \leq 1$. If $V_{\text{strat}} = 0$, the studentized statistic converges to zero, so neither tail can reject with positive limiting probability. For each nonomitted test, the population balance at the realized endpoint is nonpositive on the left and nonnegative on the right. Thus the one-sided error bounds and union bound in the proof

of Corollary 1 apply without conditioning on the grid. Outside the reporting range, boundary consistency again suffices. □

These are pointwise, fixed-type asymptotics under maintained population conditions; a distinct empirical cutoff and an empirical interior split do not verify them or establish accuracy with many sparsely observed types. Coverage for an implementation with refusal checks additionally requires refusal probability tending to zero. The stratified model targets the design-weighted observed-case mixture, not the applicant law before response or the all-randomized ITT. It supplies neither conditional randomization inference nor protection against shared shocks, and it treats the released scoring map as fixed.

S2 Inference for the Sensitivity Curve

Use a pointwise interval for worst-case movement at a prespecified loss budget, and a simultaneous band if the budget may be chosen after inspecting the curve. To assess a regression conclusion, compare movement with its margin and account for their joint uncertainty.

S2.1 Setup and pointwise assumptions

For the reference regression $y = x'\beta + u$ and a fixed contrast $a \neq 0$, recall $Q = \mathbb{E}[xx']$, $r_a = Q^{-1}a$ and $\mu_Z(a) = a'r_a$ from Section B. The directional pair is $Y(a) = (r_a'x)u$ and $Z(a) = (r_a'x)^2/\mu_Z(a)$, with $\mathbb{E}[Z(a)] = 1$. At a prespecified budget $p \in (0,1)$, write $\tau = 1 - p$ and suppress the budget argument where convenient. Define the root score, cutoff and upper-tail set by

$$W_a^* := Y(a) + \delta(p)\, Z(a), \qquad \xi_a := F_{W_a^*}^{-1}(1-p), \qquad A^* := \{W_a^* \geq \xi_a\}. \tag{S1}$$

Write the deleted and retained fractions of identifying variation and the score with known regression inputs as

$$\begin{aligned} \nu_{a,Z}^* &:= \mathbb{E}[Z(a)\mathbf{1}\{A^*\}], \\ B_a &:= 1 - \nu_{a,Z}^* = \mathbb{E}[Z(a)\mathbf{1}\{A^{*,c}\}], \\ \psi_a^* &:= (W_a^* - \xi_a)_+ + p\,\xi_a - \delta(p)\, Z(a). \end{aligned} \tag{S2}$$

For the selected moments $m_a := \mathbb{E}[x(r_a'x)\mathbf{1}\{A^*\}]$ and $m_{a,u} := \mathbb{E}[xu\mathbf{1}\{A^*\}]$, set $\delta^{\mathrm{sc}} = \delta(p)/\mu_Z(a)$ and define the direction coefficient

$$G_a := \delta^{\mathrm{sc}}\, a + \frac{1}{B_a}\Big[m_{a,u} + 2\delta^{\mathrm{sc}}(m_a - a)\Big], \qquad \zeta_a := Q^{-1}G_a. \tag{S3}$$

Write $\psi^*_{a,i}(p)$ for the evaluation of $\psi^*_a(p)$ at (x_i, y_i). The complete score accounting for estimation of the residuals and the contrast direction is

$$\Psi_{a,i}(p) = \underbrace{\psi^*_{a,i}(p)}_{\text{tail fluctuation}} - \underbrace{m_a(p)'Q^{-1}x_iu_i}_{\text{residual estimation}} - \underbrace{B_a(p)(r_a'x_i)(\zeta_a(p)'x_i)}_{\text{direction estimation}}. \tag{S4}$$

Unlike the scalar interval in Theorem 2, the curve interval uses local cutoff density bounds.

Assumption S1 (Pointwise CLT regularity)**.** *In addition to Assumptions 2 and 3, impose conditions C1–C4 below. The positive upper-tail numerator Assumption 4 is not needed for the nondegenerate directional pair (Lemma 1).*

(C1) ***Design and score envelope.*** *The moments and design of Assumption 2 and the regression setup in Section B hold, with $a \neq 0$ fixed and $0 < \mathbb{E}u^2 < \infty$; independence of x and u is not required.*

(C2) ***Quantile regularity.*** *The score $W = Y + \delta Z$ is absolutely continuous on a neighborhood of $\xi = F_W^{-1}(\tau)$, with a density f_W that is continuous and positive at ξ; in the directional case, the same local condition holds for $W(\lambda)$ at $\xi(\lambda) = F_{W(\lambda)}^{-1}(\tau)$ for λ near δ, ensuring uniqueness of the cutoff and exact upper-tail mass $P\{W(\lambda) \geq \xi(\lambda)\} = p$ there. The proofs use only the local mass bounds implied by this density condition: for small $t > 0$, each interval $(\xi - t, \xi]$ and $(\xi, \xi + t]$ has probability between positive constant multiples of t.*

(C3) ***Nonzero denominator slope.*** *With $\nu_p^Z := \mathbb{E}[Z\mathbf{1}\{W \geq \xi\}]$, the denominator slope satisfies $B := m_Z - \nu_p^Z = \mathbb{E}[Z\mathbf{1}\{W < \xi\}] > 0$ (directionally $B_a > 0$).*

(C4) ***Non-degenerate influence variance.*** $\mathrm{Var}(\Psi_a) > 0$.

The expansions use C1–C3, and nondegenerate pointwise inference also uses C4. Global threshold regularity in Assumption 5 is not required: Proposition 5 gives the fractional root without it, and C2 ensures attainment by an upper-tail region of mass

p at the root. Section S4 gives sufficient conditions for the pointwise and uniform requirements.

S2.2 Pointwise expansion and confidence interval

Theorem S1 (Expansion of the fitted criterion)**.** *Under Assumption S1 at a fixed p,*

$$\sqrt{N}\{\hat{F}_N(\delta(p), p) - F(\delta(p), p)\} = \frac{1}{\sqrt{N}} \sum_i \Psi_{a,i}(p) + o_p(1),$$

where $\hat{F}_N$ is the fractional criterion in (24)*.*

Corollary S2 (Expansion of the fitted curve)**.** *Under the same conditions,*

$$\sqrt{N}\{\hat{\delta}(p) - \delta(p)\} = \frac{1}{B_a(p)\sqrt{N}} \sum_i \Psi_{a,i}(p) + o_p(1), \qquad \omega_a^2(p) = \frac{\mathrm{Var}\{\Psi_a(p)\}}{B_a(p)^2}.$$

The limit is $\mathrm{N}(0, \omega_a^2(p))$*. Dividing by retained identifying variation $B_a(p)$ converts criterion fluctuations into movement fluctuations.*

For implementation, let $\hat{Q}_N = N^{-1}\sum_i x_i x_i'$, $\hat{r}_a = \hat{Q}_N^{-1} a$, $\hat{\mu}_Z(a) = a'\hat{r}_a$ and $\hat{u}_i = y_i - x_i'\hat{\beta}$, where $\hat{\beta}$ is OLS. Set $k = \lfloor Np \rfloor$ and $p_N = k/N$. By Proposition 4, a top-k set of fitted modified scores at the count root returns the optimal ratio in (17). Denote this set by S^*, breaking boundary ties by observation index, and let $\hat{\xi}_a(p) = \hat{W}^*_{[k]}$ be the k-th largest fitted modified score $\hat{W}_i^* = \hat{s}_i(a) + \hat{\delta}(p)\hat{Z}_i(a)$. Set $\hat{t}_{a,i} = \hat{r}_a' x_i$, $\delta^{\mathrm{s\hat{c}}} = \hat{\delta}(p)/\hat{\mu}_Z(a)$ and $\hat{B}_a = N^{-1}\sum_{i\notin S^*} \hat{Z}_i(a)$. Averaging the selected moments over all N observations gives $\hat{m}_a = N^{-1}\sum_{i\in S^*} x_i \hat{t}_{a,i}$ and $\hat{m}_{a,u} = N^{-1}\sum_{i\in S^*} x_i \hat{u}_i$, and hence

$$\hat{G}_a = \delta^{\mathrm{s\hat{c}}}\, a + \hat{B}_a^{-1}\big[\hat{m}_{a,u} + 2\delta^{\mathrm{s\hat{c}}}(\hat{m}_a - a)\big], \qquad \hat{\zeta}_a = \hat{Q}_N^{-1}\hat{G}_a. \tag{S5}$$

The fitted influence score is

$$\hat{\Psi}_{a,i} = (\hat{W}_i^* - \hat{\xi}_a)_+ + p_N\hat{\xi}_a - \hat{\delta}(p)\hat{Z}_i(a) - (x_i'\hat{Q}_N^{-1}\hat{m}_a)\hat{u}_i - \hat{B}_a\,\hat{t}_{a,i}(\hat{\zeta}_a' x_i). \tag{S6}$$

Here p_N matches the fitted top-k criterion; using p instead adds a common offset to the scores and leaves their sample variance unchanged. The standard error divides these fitted scores by retained identifying variation:

$$\widehat{\mathrm{SE}}(\hat{\delta}(p)) = \{N^{-1}\hat{B}_a(p)^{-2}\widehat{\mathrm{Var}}(\hat{\Psi}_a(p))\}^{1/2}, \tag{S7}$$

where $\widehat{\mathrm{Var}}$ is the centered empirical second moment, with divisor N.

Corollary S3 (Pointwise confidence interval)**.** *Under Assumption S1 at a prespecified budget p, the Wald interval*

$$\mathrm{CI}_{1-\alpha} = \left[\hat{\delta}(p) \pm z_{1-\alpha/2}\,\widehat{\mathrm{SE}}(\hat{\delta}(p))\right]$$

has asymptotic coverage $1-\alpha$ for $\delta(p)$.

S2.3 Estimated margins, orientation and gap inference

A known floor does not make the margin known: if b is fixed, $\hat{c} = \hat{\vartheta} - b$ still varies with the estimated coefficient (Section 2).

Assumption S2 (Asymptotically linear margin)**.** *For a population margin $c > 0$ and i.i.d. observations $O_i = (x_i, y_i)$, the estimator $\hat{c}$ satisfies*

$$\begin{aligned}\sqrt{N}(\hat{c} - c) = \frac{1}{\sqrt{N}}\sum_i \varphi_c(O_i) + o_p(1), \qquad \mathbb{E}\varphi_c = 0, \quad \mathbb{E}\varphi_c^2 < \infty,\\ \frac{1}{N}\sum_i \{\hat{\varphi}_{c,i} - \varphi_c(O_i)\}^2 = o_p(1).\end{aligned} \tag{S8}$$

Contrast margins and empirical orientation. For a prespecified contrast $a_0'\beta$ and fixed floor $b \neq a_0'\beta$, let $\iota = \mathrm{sign}(a_0'\beta - b)$ orient the adverse direction as $a = \iota a_0$. The population margin, its plug-in estimate, and its influence score are

$$c(P) = |a_0'\beta - b|, \qquad \hat{c} = |a_0'\hat{\beta} - b|, \qquad \varphi_c(x,y) = \iota(a_0'Q^{-1}x)u = (r_a'x)u.$$

In sample computation, use $\hat{\iota} = \mathrm{sign}(a_0'\hat{\beta} - b)$, $\hat{a} = \hat{\iota}a_0$, and $\hat{\varphi}_{c,i} = (\hat{r}_{\hat{a}}'x_i)\hat{u}_i$, evaluating the fitted curve and its scores in direction $\hat{a}$. For $c(P) > 0$ and consistent OLS, empirical orientation has the same asymptotic distribution as the fixed-direction estimator in direction a. A zero population margin ($a_0'\beta = b$) lies outside this regular regime: the absolute-value map is not differentiable there, and the estimated sign need not converge to a fixed orientation.

Relative floors. For an initially positive contrast and a prespecified fraction $0 \le \kappa < 1$, set $b = \kappa\vartheta$ and $c = (1-\kappa)\vartheta$. Then $\varphi_b = \kappa(r_a'x)u$ and $\varphi_c = (1-\kappa)(r_a'x)u$; empirical

substitution gives the fitted scores in (S8). Half-effect preservation uses $\kappa = 1/2$, while sign preservation uses $\kappa = 0$.

Pointwise gap interval. The gap $g(p) = \delta(p) - c$ measures how far worst-case movement exceeds the margin; estimate it by $\hat{g}(p) = \hat{\delta}(p) - \hat{c}$. Subtract the margin's influence from the curve's influence before computing the variance:

$$\varphi_{g,i}(p) = \frac{\Psi_{a,i}(p)}{B_a(p)} - \varphi_c(O_i), \qquad \hat{\varphi}_{g,i}(p) = \frac{\hat{\Psi}_{a,i}(p)}{\hat{B}_a(p)} - \hat{\varphi}_{c,i}, \tag{S9}$$

where both margin scores are zero if the margin is known. Let $\sigma_g^2(p) = \mathrm{Var}(\varphi_g(p))$ and $\hat{\sigma}_g^2(p) = \widehat{\mathrm{Var}}(\hat{\varphi}_g(p))$. Under Assumptions S1 and S2 at a prespecified p with $\sigma_g^2(p) > 0$, the Wald interval with asymptotic coverage $1 - \alpha$ is

$$\left[\hat{g}(p) \pm z_{1-\alpha/2}\hat{\sigma}_g(p)/\sqrt{N}\right],$$

which accounts for the covariance between movement and the estimated margin.

S2.4 Simultaneous bands and implementation

For simultaneous bands, take $\mathscr{P} = [\underline{p}, \overline{p}] \subset (0,1)$ compact and strengthen the pointwise conditions as follows.

Assumption S3 (Uniform regularity in p)**.** *Assumption S1 holds at every $p \in \mathscr{P}$, together with conditions U1–U2 below.*

(U1) ***Uniform quantile regularity.*** *There exist an open neighborhood $\mathscr{U}_0 \subset \mathbb{R} \times (0,1)$ of the root path $\{(\delta(p), p) : p \in \mathscr{P}\}$ and constants $\eta_0 > 0$ and $0 < c_f \leq C_f < \infty$ such that for every $(\lambda, p) \in \mathscr{U}_0$, the score $W(\lambda) = Y(a) + \lambda Z(a)$ has a continuous density with values in $[c_f, C_f]$ on $(\xi(\lambda,p) - \eta_0, \xi(\lambda,p) + \eta_0)$, where $\xi(\lambda,p)$ is its $(1-p)$-quantile. Consequently, the local probability bound $\sup_{(\lambda,p)\in\mathscr{U}_0} P\left(|W(\lambda) - \xi(\lambda,p)| \leq t\right) \leq 2C_f t$ holds for all small $t > 0$.*

(U2) ***Uniformly nonzero slope and variance.*** *$\inf_{p\in\mathscr{P}} B_a(p) > 0$ and $\inf_{p\in\mathscr{P}} \mathrm{Var}(\Psi_a(p)) > 0$, the uniform counterparts of C3 and C4.*

Under U1 and C1, the first-stage coefficients $p \mapsto m_a(p)$ and $p \mapsto \zeta_a(p)$ are continuous on $\mathscr{P}$ by dominated convergence and cutoff atomlessness.

Lemma S1 (The sensitivity curve)**.** *Under Assumption S3, the map $p \mapsto \delta(p)$ is continuous and strictly increasing on $\mathscr{P}$. On the interior of $\mathscr{P}$ it is continuously differentiable, with*

$$\partial_p \delta(p) = \frac{\xi_a(p)}{B_a(p)}. \tag{S10}$$

Theorem S2 (Functional central limit theorem)**.** *Under Assumption S3, the expansions in Theorem S1 and corollary S2 hold uniformly over $p \in \mathscr{P}$. In $\ell^\infty(\mathscr{P})$, the process $\sqrt{N}(\hat{\delta} - \delta)$ converges weakly to a tight mean-zero Gaussian process $\mathbb{G}$ with almost surely uniformly continuous paths and covariance kernel*

$$K(p,q) = \frac{\mathbb{E}[\Psi_a(p)\Psi_a(q)]}{B_a(p)B_a(q)}, \qquad K(p,p) = \omega_a^2(p). \tag{S11}$$

S2.4.1 Band implementation with paired scores

Using independent standard-normal multipliers $e_1, \dots, e_N$, independent of the data and shared by the curve and margin across all $p \in \mathscr{P}$, form

$$\hat{\mathbb{G}}^*_{g,N}(p) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_i \{\hat{\varphi}_{g,i}(p) - \bar{\hat{\varphi}}_g(p)\}, \qquad \bar{\hat{\varphi}}_g(p) = \frac{1}{N} \sum_{i=1}^{N} \hat{\varphi}_{g,i}(p),$$

and let $\hat{q}^g_{1-\alpha}$ be the conditional $(1-\alpha)$-quantile of $\sup_{p\in\mathscr{P}} |\hat{\mathbb{G}}^*_{g,N}(p)|/\hat{\sigma}_g(p)$. Sharing multipliers preserves the estimated margin's covariance with the curve.

Theorem S3 (Functional limit and simultaneous band)**.** *Under Assumptions S2 and S3 and $\inf_{p\in\mathscr{P}} \sigma_g^2(p) > 0$,*

$$\mathbb{P}\Big(g(p) \in \big[\hat{g}(p) \pm \hat{q}^g_{1-\alpha} \hat{\sigma}_g(p)/\sqrt{N}\big] \ \forall p \in \mathscr{P} \Big) \longrightarrow 1 - \alpha.$$

The process $\sqrt{N}(\hat{g} - g)$ also converges weakly in $\ell^\infty(\mathscr{P})$ to the tight Gaussian process with covariance $\mathbb{E}[\varphi_g(p;O)\varphi_g(q;O)]$.

Setting both margin scores to zero gives a known-margin gap band; adding back c gives the curve band. An estimated margin requires a positive lower bound on gap variance, not just on curve variance. With fixed tie-breaking, the fitted gap scores and their standard deviation depend on p only through $k = \lfloor Np \rfloor$. The studentized multiplier statistic is therefore constant on each count cell $[k/N, (k+1)/N)$, so evaluating every represented count gives its exact supremum over $\mathscr{P}$. Section S5.3 compares coverage at count-cell starts with coverage over the full reporting range.

S2.5 Connection to scalar tipping inference

For margins c in the image $\delta(\operatorname{int}\mathscr{P})$, the unclipped tipping fraction is $p^\dagger(c) = \delta^{-1}(c)$, with $\frac{d}{dc}p^\dagger(c) = B_a(p^\dagger(c))/\xi_a(p^\dagger(c)) > 0$. With orientation fixed, raising the floor b can only enlarge the class of exclusions that overturn the conclusion, so the population tipping fraction is nonincreasing in b by Definition 1. Reported interval endpoints may move in either direction.

At the scalar crossing $p = p^\dagger$, suppose Proposition 2 applies and the ratio conditions Assumptions S1 and S2 hold, with $\sigma_g^2(p) > 0$. For the coefficient on w, use the same orientation and floor in both procedures, so $c = \vartheta - b$ and $\hat{c} = \hat{\vartheta} - \hat{b}$ for the floor estimate used in $\hat{A}_b$. Since $\delta(p^\dagger) = c$ (Proposition 6), the two scores select the same upper tail, with normalization

$$q_b = L(Y + cZ), \qquad \xi_* = L\xi_a(p^\dagger), \qquad B_v(p^\dagger) = LB_a(p^\dagger),$$

where B_v is the retained identifying variation in Section A and B_a is its fraction of the reference total L. The complete scores satisfy

$$\varphi_A(p) = L\{\Psi_a(p) - B_a(p)\varphi_c\} = LB_a(p)\varphi_g(p). \tag{S12}$$

Consistency of the standard deviations and (S12) imply that $\sqrt{N}\,\hat{A}_b(p)/\hat{\sigma}_A(p)$ and $\sqrt{N}\,\hat{g}(p)/\hat{\sigma}_g(p)$ differ by $o_p(1)$ at the crossing. At an interior crossing, the cutoff scaling $\xi_* = L\xi_a(p)$ cancels the factor L in (S12), so the tipping variance in (6) equals $\operatorname{Var}(\Psi_a(p) - B_a(p)\varphi_c)/\xi_a(p)^2$.

S3 Proofs for Curve Inference

S3.1 Local expansions and selection bounds

Use the upper-tail set and scores in (S1)–(S4), suppressing the budget argument where convenient. The selection bounds handle estimated scores without differentiating the selected set. For the direction correction, define the scalar root $\delta^{\mathrm{sc}}(r)$ for $a'r > 0$ by

$$f(\varrho; r) := \mathrm{T}_p\{(r'x)u + \varrho(r'x)^2\} - \varrho(r'Qr) = 0, \tag{S13}$$

which is the population Dinkelbach criterion of Proposition 5 evaluated at the unnormalized pair $((r'x)u, (r'x)^2)$ for direction r. Here ϱ is the multiplier on that unnormalized scale, not the normalized movement λ of (21). Define the associated directional functional

$$\Phi_a(r) := (a'r)\,\delta^{\mathrm{sc}}(r), \qquad \Phi_a(r_a) = \delta, \;\; \delta^{\mathrm{sc}}(r_a) =: \delta^{\mathrm{sc}} = \delta/\mu_Z(a). \tag{S14}$$

Positive homogeneity of T_p makes Φ_a homogeneous of degree zero. We first identify the coefficient in (S3) as the gradient of Φ_a.

The standing moments give $H := 1 + \|x\|\,|u| + \|x\|^2 \in L^2(P)$, since $\mathbb{E}[\|x\|^2 u^2] = \mathbb{E}[\|x\|^2 \sigma^2(x)] < \infty$ and $\mathbb{E}\,\|x\|^4 < \infty$. Write $\hat\theta := (\hat\beta, \hat r_a)$, $\theta_0 := (\beta, r_a)$ and $\hat Q_N := N^{-1}\sum_i x_i x_i'$. The OLS normal equation and the inverse-matrix identity yield the following expansions using $\mathbb{E}[xu] = 0$ and $\hat Q_N \xrightarrow{p} Q \succ 0$:

$$\begin{aligned} \left\|\hat\theta - \theta_0\right\| &= O_p(N^{-1/2}), \\ \sqrt{N}(\hat\beta - \beta) &= Q^{-1}N^{-1/2}\sum_i x_i u_i + o_p(1), \\ \sqrt{N}(\hat r_a - r_a) &= -Q^{-1}N^{-1/2}\sum_i (x_i x_i' - Q) r_a + o_p(1). \end{aligned} \tag{S15}$$

For a candidate coefficient vector β° and direction r, write $\theta = (\beta^\circ, r)$, $t_a = r_a'x$ and $c_0 = a'r_a = \mu_Z(a)$. On compact neighborhoods with $a'r \geq c_0/2$, define

$$Y_\theta = (r'x)(y - x'\beta^\circ), \qquad Z_\theta = \frac{(r'x)^2}{a'r}, \qquad V_\theta(\lambda) = Y_\theta + \lambda Z_\theta. \tag{S16}$$

At θ_0 their gradients are column vectors

$$D_{\beta^\circ} Y_{\theta_0} = -t_a x, \qquad D_r Y_{\theta_0} = xu, \qquad D_r Z_{\theta_0} = \frac{2t_a x}{c_0} - \frac{t_a^2 a}{c_0^2}.$$

The nuisance coefficients of the criterion at the population root, including the full-sample subtraction, are

$$\begin{aligned} \mathbb{E}[D_{\beta^\circ} Y_{\theta_0} \mathbf{1}\{A^*\}] &= -m_a, \\ \mathsf{D}_a &:= \mathbb{E}[(D_r Y_{\theta_0} + \delta D_r Z_{\theta_0})\mathbf{1}\{A^*\}] - \delta\mathbb{E}[D_r Z_{\theta_0}] \\ &= m_{a,u} + 2\delta^{\mathrm{sc}} m_a - \delta^{\mathrm{sc}}(\nu^*_{a,Z} + 1)a, \qquad \delta^{\mathrm{sc}} = \delta/c_0. \end{aligned} \tag{S17}$$

Here $\mathbb{E}[D_r Z_{\theta_0}] = a/c_0$; since $B_a = 1 - \nu^*_{a,Z}$, the ratio D_a/B_a gives the expression in (S3). Its identification as the population gradient is established below, separately from this algebra.

Lemma S2 (Direction differentiability)**.** *Under the standing conditions and* C1–C3*, the population functional* Φ_a *in* (S14) *is differentiable at* r_a*, with* $G_a = \mathsf{D}_a/B_a$*.*

Proof. Fix $\beta^\circ = \beta$, let $q = (r, \lambda)$, $q_0 = (r_a, \delta)$, and write $V_q = Y_{(\beta,r)} + \lambda Z_{(\beta,r)}$, $Z_r = Z_{(\beta,r)}$, and $A_0 = \{V_{q_0} \geq \xi_a\}$. Use the tail-value representation $\mathrm{T}_p(V) = \sup_{0\leq w\leq 1,\, \mathbb{E}w=p} \mathbb{E}[Vw]$; fractional cutoff weights accommodate neighboring atoms without changing the value. At the base point, the maximizing weight is uniquely $w_0 = \mathbf{1}\{A_0\}$ almost surely. All neighboring cutoffs converge to ξ_a by the strict population count gaps at $\xi_a \pm \eta$ and L^1 convergence of the scores. Since $|V_q - V_{q_0}| \leq C \,\|q - q_0\|\, H$, every maximizing weight w_q differs from w_0 only in a shrinking base-cutoff band or where H is large. Truncation and cutoff atomlessness give $\mathbb{E}[H|w_q - w_0|] \to 0$. Thus the population sandwich

$$\mathbb{E}[(V_q - V_{q_0})w_0] \leq \mathrm{T}_p(V_q) - \mathrm{T}_p(V_{q_0}) \leq \mathbb{E}[(V_q - V_{q_0})w_q]$$

has gap $o(\|q - q_0\|)$. The score map has an L^1 Taylor expansion, so the population criterion $\mathrm{T}_p(V_q) - \lambda\mathbb{E}[Z_r]$ is differentiable at the base point with partial derivatives $-B_a$ in λ and D_a in r. The standing denominator assumption gives $B_{\min} = \int_0^\tau F^{-1}_{Z_{r_a}}(s)\, ds > 0$. Every nearby retained weight of mass τ has leverage at least $B_{\min} - \mathbb{E}|Z_r - Z_{r_a}| \geq B_{\min}/2$. Hence the population criterion has slope at most $-B_{\min}/2$ in λ; continuity preserves the opposite signs at two fixed points bracketing δ, giving a unique nearby root $\lambda(r)$. Evaluating the criterion at the base root first gives $|\lambda(r) - \delta| = O(\|r - r_a\|)$. The joint differentiability just proved then yields

$$\lambda(r) - \delta = {B_a}^{-1}\mathsf{D}'_a(r - r_a) + o(\|r - r_a\|).$$

Rescaling the normalized root equation by $\varrho = \lambda/(a'r)$ gives exactly (S13); thus $\lambda(r) = \Phi_a(r)$ and $\nabla\Phi_a(r_a) = \mathsf{D}_a/B_a$. □

Proposition S2 (Uniform convergence and consistency)**.** *Under Assumptions 2 to 4, for every compact* $[0, \Lambda]$*,* $\sup_{\lambda\in[0,\Lambda]} |f_N(\lambda) - f(\lambda)| \xrightarrow{p} 0$ *and* $\hat\delta \xrightarrow{p} \delta$*, the population root for fractional exclusions. For the directional pair under the standing regression moments, this gives* $\hat\delta \xrightarrow{p} \delta$ *without a cutoff regularity condition.*

Proof. Uniform convergence. For $H_0 := |Y| + \Lambda Z$, choose K such that $P(H_0 > K) < \min(p, \tau)$. Since $|W(\lambda)| \le H_0$, the population cutoff inequalities place every minimizer in $[-K, K]$. With probability tending to one, $P_N \mathbf{1}\{H_0 > K\} < \min(p_N, 1 - p_N)$; the count inequalities in Lemma 4 then place every empirical minimizer in the same interval, uniformly in λ and including ties. On this compact parameter rectangle, the criterion class is VC-subgraph with integrable envelope proportional to $H_0 + K$. The uniform law of large numbers, with rounding error bounded by $K|p_N - p|$, gives uniform criterion convergence before and hence after minimizing over ξ.

Consistency. The positive root in Proposition 5 is bracketed by $f(\delta - \eta) > 0 > f(\delta + \eta)$ for sufficiently small $\eta > 0$. Convergence at those endpoints and $\mathbb{P}(\mathscr{D}_N) \to 1$ yield the same bracket for $\hat{\delta}$ with probability tending to one.

For the fitted directional statistic, OLS consistency and the top-k contraction give

$$\sup_{\lambda \in [0,\Lambda]} |f_{N,\hat{\theta}}(\lambda) - f_{N,\theta_0}(\lambda)| \le C_\Lambda \left\| \hat{\theta} - \theta_0 \right\| P_N H = o_p(1),$$

where $f_{N,\theta}$ uses the pair in (S16). The fitted minimum retained denominator differs from the bottom-$(N-k)$ oracle leverage average by at most $P_N |Z_{\hat{\theta}} - Z(a)| = o_p(1)$, so the same sign brackets apply to the fitted criterion. □

S3.1.1 Value and selection bounds

Write $k(p) = \lfloor pN \rfloor$, $p_N(p) = k(p)/N$ and P_N for the empirical average. For the pair in (S16), define the count criterion and its population counterpart by

$$f_{N,\theta}(\lambda; p) = T_N(V_\theta(\lambda)) - \lambda P_N Z_\theta, \qquad T_N(W) = \max_{|S| = k(p)} P_N[W \mathbf{1}\{S\}].$$

At $\hat{\theta}$ the count root is $\hat{\delta}(p)$ and $P_N Z_{\hat{\theta}} = 1$. The population criterion is F in (21).

Lemma S3 (Value and root stability)**.** *For $1 \le k < N$ and arrays V, W with cardinality-k top sets S_V, S_W,*

$$P_N[(V-W)\mathbf{1}\{S_W\}] \le T_N(V) - T_N(W) \le P_N[(V-W)\mathbf{1}\{S_V\}], \quad |T_N(V) - T_N(W)| \le P_N|V-W|.$$

For $f(\lambda) = T_N(Y + \lambda Z) - \lambda P_N Z$, $Z \ge 0$, and top sets S_j at λ_j, $\lambda_1 < \lambda_2$,

$$-P_N[Z\mathbf{1}\{S_1^c\}] \le \frac{f(\lambda_2) - f(\lambda_1)}{\lambda_2 - \lambda_1} \le -P_N[Z\mathbf{1}\{S_2^c\}].$$

If every retained denominator is at least $B_{\min} > 0$, then f has a unique root ℓ, and for any comparison point ℓ_0,

$$|\ell - \ell_0| \le |f(\ell_0)|/B_{\min}.$$

For two leverage arrays, uniformly over $|S| = k$,

$$P_N[Z_\theta \mathbf{1}\{S^c\}] \ge N^{-1} \sum_{j=k+1}^{N} Z_{[j]}(\theta_0) - P_N|Z_\theta - Z_{\theta_0}|.$$

Proof. Evaluate each maximum at the other maximizing set to obtain the value sandwich. Substitute $V = Y + \lambda_2 Z$ and $W = Y + \lambda_1 Z$ and subtract $(\lambda_2 - \lambda_1)P_N Z$ to obtain the slope bounds. A positive denominator floor gives the unique ratio root by Proposition 4, and applying the slope bound between that root and ℓ_0 gives root stability. The last inequality follows by replacing Z_θ with Z_{θ_0} on S^c and minimizing the latter retained sum. □

Lemma S4 (Cutoff trapping and weighted selection)**.** *Let $O_i = (x_i, y_i)$ denote observation i and $W_t(O_i)$ its oracle score. Index scores and cutoffs by $t \in K$, with $A_t = \{W_t \ge \xi_t\}$, $P(A_t) = p_t$ and $k_N(t)/N = p_t + O(N^{-1})$ uniformly. Suppose their level and band indicators obey uniform laws of large numbers and, for each small $\eta > 0$,*

$$\inf_t \min\{p_t - P(W_t \ge \xi_t + \eta),\ P(W_t > \xi_t - \eta) - p_t\} > 0, \qquad \sup_t P(|W_t - \xi_t| \le \eta) \longrightarrow 0$$

as $\eta \downarrow 0$ in the second expression. Let $H \ge 1$, $PH^2 < \infty$ and $\varepsilon_N \downarrow 0$ be deterministic. Uniformly over all perturbed arrays with $|V_i - W_t(O_i)| \le C\varepsilon_N H_i$, for every measurable cardinality-$k_N(t)$ top set S and every threshold minimizer $\hat{\xi}$ in Lemma 4,

$$\sup |\hat{\xi} - \xi_t| \xrightarrow{p} 0, \qquad \sup P_N[L\mathbf{1}\{S \triangle A_t\}] \xrightarrow{p} 0, \quad L \in \{1, H, H^2\}.$$

Proof. For fixed $\eta > 0$, the empirical fraction with $|V_i - W_t(O_i)| > \eta/2$ is at most $2C\varepsilon_N P_N H/\eta = o_p(1)$. The oracle count gaps at $\xi_t \pm \eta/2$, their uniform laws, and budget rounding therefore give fewer than $k_N(t)$ scores at or above $\xi_t + \eta$ and more than $k_N(t)$ strictly above $\xi_t - \eta$. These inequalities trap both adjacent order statistics and hence the entire threshold-minimizer interval uniformly. Membership in a top set then gives

$$S \triangle A_t \subseteq \{|W_t - \xi_t| \le \eta + C\varepsilon_N H\},$$

irrespective of tie allocation. For fixed M and eventually $C\varepsilon_N M \le \eta$, its weighted mass is at most

$$M^2 P_N\{|W_t - \xi_t| \le 2\eta\} + P_N[H^2 \mathbf{1}\{H > M\}].$$

First let $N \to \infty$ and then $\eta \downarrow 0$ for fixed M; the uniform law for the bands and the vanishing band probabilities eliminate the first term. Finally let $M \to \infty$; integrability of H^2 eliminates the remaining tail term. □

For the next bounds, let $\mathscr{I}$ be a singleton $\{p\}$ satisfying C1–C3 or a compact interval $\mathscr{P}$ satisfying U1–U2 in addition to the pointwise conditions. The standing assumptions apply on $\mathscr{I}$. Write A_p^*, $\xi_a(p)$ and $B_a(p)$ for the quantities in (S1)–(S2) at budget p.

Lemma S5 (Criterion and retained-leverage bounds)**.** *For each deterministic $\varepsilon_N \downarrow 0$, the following bounds hold uniformly over $p \in \mathscr{I}$ and the indicated parameter ranges.*

(a) *For θ, θ' in a compact neighborhood with $a'r, a'r' \ge c_0/2$ and bounded λ, $|f_{N,\theta}(\lambda; p) - f_{N,\theta'}(\lambda; p)| \le C \|\theta - \theta'\| P_N H$.*

(b) *For $\|\theta - \theta_0\| \le \varepsilon_N$ and $|\lambda - \delta(p)| \le \varepsilon_N$, the threshold minimizers converge to $\xi_a(p)$ and every cardinality-$k(p)$ top set S satisfies*

$$P_N[H^2 \mathbf{1}\{S \triangle A_p^*\}] = o_p(1), \qquad P_N[Z_\theta \mathbf{1}\{S^c\}] = B_a(p) + o_p(1).$$

(c) *At $\lambda = \delta(p)$ and for $\|\theta - \theta_0\| \le \varepsilon_N$,*

$$f_{N,\theta}(\delta(p); p) - f_{N,\theta_0}(\delta(p); p) = -m_a(p)'(\beta^\circ - \beta) + \mathsf{D}_a(p)'(r - r_a) + o_p(1)\|\theta - \theta_0\| + O_p(1)\|\theta - \theta_0\|^2.$$

(d) *With $B_{\min} = \int_0^{1-\max \mathscr{I}} F^{-1}_{Z(a)}(s)\, ds > 0$, every size-$k(p)$ set has retained leverage at least $B_{\min}/2$ with probability tending to one, for all $\|\theta - \theta_0\| \le \varepsilon_N$.*

Proof. The product and quotient bounds for (S16) give

$$|V_\theta(\lambda) - V_{\theta'}(\lambda)| \le C(1 + |\lambda|)\|\theta - \theta'\| H, \qquad |Z_\theta - Z_{\theta'}| \le C\|\theta - \theta'\| \|x\|^2.$$

Apply the contraction in Lemma S3 for part (a).

For part (b), apply Lemma S4 to $(W_a^*(p), \xi_a(p), p)$. The root path is bounded by $\mathbb{E}|Y(a)|/B_{\min}$ by population retained leverage, so the score perturbation is bounded

by $C\varepsilon_N H$. Condition C2 at a singleton and U1 on an interval give the required count gaps and vanishing band probabilities. The level sets and bands are half-planes and intersections of half-planes in $(Y(a), Z(a))$, so their indicator classes are VC. Their uniform laws extend to multiplication by fixed integrable weights via bounded simple-function approximation and tail truncation. In particular,

$$|P_N[Z_\theta \mathbf{1}\{S^c\}] - B_a(p)| \le P_N|Z_\theta - Z(a)| + P_N[Z(a)\mathbf{1}\{S \triangle A_p^*\}] + |(P_N - P)[Z(a)\mathbf{1}\{A_p^{*,c}\}]| = o_p(1)$$

uniformly, using $Z(a) \le CH$.

For part (c), put $\Delta\theta = (\Delta\beta^\circ, \Delta r) = \theta - \theta_0$ and $\lambda = \delta(p)$. The observation-wise Taylor expansion is

$$V_\theta(\lambda) - V_{\theta_0}(\lambda) = (D_{\beta^\circ} Y_{\theta_0})'\Delta\beta^\circ + (D_r Y_{\theta_0} + \lambda D_r Z_{\theta_0})'\Delta r + R_\theta, \qquad |R_\theta| \le CH \left\|\Delta\theta\right\|^2.$$

Replace both bounding sets in the value sandwich by A_p^* using part (b), at cost $o_p(1)\left\|\Delta\theta\right\|$, and then apply the weighted uniform laws to the derivative coordinates. The remainder averages to $O_p(1)\left\|\Delta\theta\right\|^2$. Expanding $-\lambda P_N Z_\theta$ as well yields the direction coefficient in (S17).

For part (d), use the bottom-trimmed-sum bound in Lemma S3 at $k(\max \mathscr{F})$. The normalized bottom sum converges to $B_{\min}$ by the trimmed-sum law, and the perturbation is bounded by $C\varepsilon_N P_N \|x\|^2 = o_p(1)$. □

S3.2 Influence functions and pointwise inference

Theorem S4 (Design influence function)**.** *Under Assumption S1, the coefficient G_a in (S3) equals $\nabla_r \Phi_a(r)|_{r=r_a}$, and $\hat{r}_a$ contributes to $\sqrt{N}(\hat{\delta} - \delta)$ the first-order term $G_a'\sqrt{N}(\hat{r}_a - r_a)$. The gradient is orthogonal to the reference contrast direction r_a, so $G_a' r_a = 0$ by Euler's relation. The per-observation design influence is*

$$d_{a,i} := G_a' \varphi_{r,i} = -(r_a' x_i)(\zeta_a' x_i), \qquad \varphi_{r,i} := r_a - (r_a' x_i) Q^{-1} x_i, \qquad \zeta_a := Q^{-1} G_a, \quad \text{(S18)}$$

where $\varphi_{r,i}$ is the influence function of $\hat{r}_a$, ζ_a re-expresses the design gradient G_a in the Q-metric, and $\mathbb{E}[d_{a,i}] = 0$. The contribution of $d_{a,i}$ to the criterion score in (S4) is $B_a d_{a,i}$.

Proof. Lemma S2 gives $G_a = \mathsf{D}_a / B_a$, and (S17) gives its closed form. For $G_a' r_a = 0$, the root identity $\mathbb{E}[W_a^* \mathbf{1}\{W_a^* \ge \xi_a\}] = \delta$ with $W_a^* = Y(a) + \delta Z(a)$ implies $r_a' m_{a,u} = \delta B_a$ and $r_a' m_a = \mu_Z(a)\nu_{a,Z}^*$; substituting and using $\delta^{\mathrm{sc}}\mu_Z(a) = \delta$ collapses $G_a' r_a$ to zero.

Finally $\varphi_{r,i} = -Q^{-1}(x_i x_i' - Q)r_a = r_a - (r_a' x_i)Q^{-1}x_i$ is the influence function of $\hat{Q}_N^{-1}a$; using $G_a' r_a = 0$ yields $G_a'\varphi_{r,i} = -(r_a' x_i)(G_a' Q^{-1} x_i) = -(r_a' x_i)(\zeta_a' x_i)$ with $\zeta_a = Q^{-1}G_a$. The criterion coefficient $B_a G_a$ becomes the root coefficient G_a by the fitted slope calculation in the proof of Corollary S2. □

Corollary S4 (Partial adaptivity). *If $G_a = 0$, direction estimation has no first-order effect: the influence numerator is $\tilde{\psi}_{a,i} := \psi^*_{a,i} - m_a' Q^{-1} x_i u_i$ and the normalized influence is $\tilde{\psi}_{a,i}/B_a$ in (S4). Index sufficiency and mean-linearity jointly imply $G_a = 0$. Both hold for mean-zero elliptically symmetric x with $u \perp\!\!\!\perp x$, and at $d = 1$.*

For the index $t = r_a' x$, let $\Delta_x(t) := \mathbb{E}[x \mid t] - \mu_Z(a)^{-1} a\, t$ be the deviation from the linear conditional mean. Under index sufficiency $\mathbb{E}[x \mid t, u] = \mathbb{E}[x \mid t]$, conditioning in (S3) gives $B_a G_a = \mathbb{E}[\Delta_x(t)(u + 2\delta^{\mathrm{sc}} t)\mathbf{1}\{W_a^* \geq \xi_a\}]$. Mean-linearity makes $\Delta_x \equiv 0$. At $d = 1$, x is determined by its index, and Euler's relation $G_a' r_a = 0$ also gives $G_a = 0$ directly.

Proof of Theorem S1 and corollary S2 and their uniform extension. All remainders below are uniform over the budget set $\mathscr{I}$ of Lemma S5, proving also the uniform expansions asserted in Theorem S2.

Hinge term at the population root. For possible values (s, ℓ) of the transformed pair $(Y(a), Z(a))$, define the hinge criterion

$$\chi_{\lambda,\xi,p}(s, \ell) = (s + \lambda\ell - \xi)_+ + p\xi - \lambda\ell. \tag{S19}$$

Positive-part contraction gives a parameter Lipschitz modulus $2\ell + 2 + \sup|\xi|$ and envelope $C(1 + |s| + \ell)$ on any compact parameter domain. Both belong to $L^2(P)$ by C1, so Lemma 2 makes the class pointwise measurable and P-Donsker. Write $\chi_\xi = \chi_{\delta(p),\xi,p}(Y(a), Z(a))$; by Lemma 4, $f_{N,\theta_0}(\delta(p); p) = \min_\xi \Gamma_N(\xi)$ with $\Gamma_N(\xi) = P_N \chi_\xi + (p_N(p) - p)\xi$. Any minimizer ξ_N^0 satisfies $\xi_N^0 - \xi_a(p) = o_p(1)$ by Lemma S5(b), and empirical and population optimality give

$$(P_N - P)(\chi_{\xi_N^0} - \chi_{\xi_a(p)}) + (p_N(p) - p)(\xi_N^0 - \xi_a(p)) \leq \Gamma_N(\xi_N^0) - \Gamma_N(\xi_a(p)) \leq 0.$$

Because $\mathscr{I}$ is bounded away from 0 and 1, localization places both cutoffs in a fixed compact interval with probability tending to one. The hinge Lipschitz bound then gives $\sup_{p\in\mathscr{I}} \left\| \chi_{\xi_N^0} - \chi_{\xi_a(p)} \right\|_{L^2(P)} = o_p(1)$. Asymptotic $L^2(P)$ equicontinuity of the Donsker class makes the first term uniformly $o_p(N^{-1/2})$, while rounding costs

$O_p(N^{-1})$. Since $\chi_{\xi_a(p)} = \psi_a^*(p)$ and $P\chi_{\xi_a(p)} = F(\delta(p), p) = 0$, the count criterion at θ_0 is $P_N\psi_a^*(p) + o_p(N^{-1/2})$.

Fitted criterion. Apply Lemma S5(c) at $\hat{\theta}$ (taking $\varepsilon_N = N^{-1/4}$), and insert (S15) and (S18) to obtain

$$f_{N,\hat{\theta}}(\delta(p); p) = P_N\Psi_a(p) + o_p(N^{-1/2}).$$

The score is centered and square integrable by the hinge identity and the OLS influence formulas. The fractional correction in (24) is $O_p(N^{-1})$ by cutoff trapping of the next fitted order statistic, proving Theorem S1.

Fitted root. The Donsker bound on the hinge term and the $\sqrt{N}$-consistent OLS rates with bounded coefficients give $f_{N,\hat{\theta}}(\delta(p); p) = O_p(N^{-1/2})$. The retained-leverage floor in Lemma S5(d) and root stability in Lemma S3 yield

$$|\hat{\delta}(p) - \delta(p)| \le 2|f_{N,\hat{\theta}}(\delta(p); p)|/B_{\min} = O_p(N^{-1/2}).$$

At both endpoints, Lemma S5(b) now gives fitted retained leverage $B_a(p) + o_p(1)$, so the secant bounds give

$$f_{N,\hat{\theta}}(\delta(p); p) = \{B_a(p) + o_p(1)\}\{\hat{\delta}(p) - \delta(p)\},$$

which also holds trivially when the roots coincide. Since $B_a(p) \ge B_{\min}$, dividing gives Corollary S2. □

S3.2.1 Variance consistency

The fitted score depends on the selected tail through finitely many coefficients, so one compact-parameter Lipschitz bound serves both pointwise variance consistency and the multiplier proof.

Lemma S6 (Empirical covering by a parameter mesh)**.** *On a fixed compact domain for* $\theta = (e, r, \lambda, \mu, \xi, p, B, t, \zeta)$*, with* $e, r, t, \zeta \in \mathbb{R}^d$ *and* μ, B *bounded away from zero, define*

$$h_\theta(x, u) = \frac{\big((r'x)(u - x'e) + \lambda(r'x)^2/\mu - \xi\big)_+ + p\xi - \lambda(r'x)^2/\mu}{B} - (x't)(u - x'e) - (r'x)(\zeta'x).$$

This family has envelope and parameter Lipschitz modulus $M = C_\Theta H \in L^2(P)$*. Its population and empirical class bounds are those of Lemma 2, with parameter count* $d_\Theta = 4d + 5$*. If fitted parameters lie uniformly in this domain with probability tending*

to one, denote their score class by $\hat{\mathcal{G}}_{\Psi,N}$. *On events* $\mathscr{C}_N$ *with probability tending to one, it has the deterministic entropy profile*

$$\log \mathcal{N}(\varepsilon, \hat{\mathcal{G}}_{\Psi,N}, \|\cdot\|_{N,2}) \le \psi(\varepsilon) := d_\Theta \log(1 + C''_\Theta/\varepsilon), \qquad 0 < \varepsilon \le 1.$$

With an enlarged constant, the same entropy bound holds in $L^2(P)$ *for the population score class.*

Proof. Positive-part contraction and the product and quotient rules give $|h_\theta| \le M$ and $|h_\theta - h_{\theta'}| \le M \|\theta - \theta'\|$. Apply Lemma 2 and intersect fitted localization with $\{\|M\|_{N,2} \le 2(PM^2)^{1/2}\}$ to obtain $\mathscr{C}_N$. The strong law makes this single event valid simultaneously at every covering radius. For the curve influence function $\Psi_a(p)/B_a(p)$ take

$$(e, r, \lambda, \mu, \xi, p, B, t, \zeta) = (0, r_a, \delta(p), \mu_Z(a), \xi_a(p), p, B_a(p), Q^{-1} m_a(p)/B_a(p), \zeta_a(p)).$$

For the fitted slope-normalized score, take the corresponding fitted quantities, with $e = \hat{\beta} - \beta$ and budget coordinate $p_N(p)$. The two appearances of λ use the same root, and all fitted coefficients use the same cardinality-$k(p)$ selection. Since $u - x'e = \hat{u}$ and division by $\hat{B}_a$ cancels the factor $\hat{B}_a$ in the direction-estimation term of (S6), this substitution reproduces $\hat{\Psi}_a(p)/\hat{B}_a(p)$ exactly, including at $d = 1$. Selected-set changes thus enter through finitely many coefficients, not an additional observation-specific indicator. □

Proposition S3 (Plug-in consistency)**.** *Let* $\mathscr{I} = \{p\}$ *under Assumption S1, or* $\mathscr{I} = \mathscr{P}$ *under Assumption S3. The fitted coefficients in* (S6) *are uniformly consistent on* $\mathscr{I}$ *and satisfy the compact localization of Lemma S6. Writing* $\Delta(p) = \Psi_a(p)/B_a(p)$ *and* $\hat{\Delta}(p) = \hat{\Psi}_a(p)/\hat{B}_a(p)$,

$$\sup_{p \in \mathscr{I}} |\hat{B}_a(p) - B_a(p)| \xrightarrow{p} 0,$$
$$\sup_{p \in \mathscr{I}} \left\{ \left\| \hat{\Psi}_a(p) - \Psi_a(p) \right\|_{N,2} + \left\| \hat{\Delta}(p) - \Delta(p) \right\|_{N,2} \right\} \xrightarrow{p} 0,$$

where $\|h\|_{N,2} = (P_N h^2)^{1/2}$. *In particular, at a fixed* p, $\sqrt{N}\,\widehat{\mathrm{SE}}(\hat{\delta}(p)) \xrightarrow{p} \omega_a(p)$.

Proof of Proposition S3. The preceding root expansion and (S15) give uniform root and OLS consistency. Lemma S5(b) gives uniform cutoff and weighted selection

consistency. Its weighted laws yield $(\hat{m}_a, \hat{m}_{a,u}, \hat{B}_a) \xrightarrow{p} (m_a, m_{a,u}, B_a)$ uniformly: the oracle weights are bounded by CH, and fitting residuals and directions costs $o_p(1)P_NH$. The slope floor, $\hat{Q}_N \xrightarrow{p} Q \succ 0$ and (S5) then give uniform consistency of $(\hat{G}_a, \hat{\xi}_a)$. On an interval, Lemma S1 and dominated convergence across null cutoff events give continuous population paths; compactness and the denominator floors then localize the fitted parameters, also at a singleton. The score Lipschitz bound in Lemma S6, $p_N - p = O(N^{-1})$ and $P_NH^2 \xrightarrow{p} \mathbb{E}H^2$ give both empirical-L^2 conclusions. At fixed p, the second-moment law and Cauchy–Schwarz give variance and standard-error consistency. □

Margin scores and empirical orientation. The OLS expansion in (S15) and the delta method give the fixed- and relative-floor margin scores in Section S2.3. The maintained moment and OLS consistency conditions deliver the empirical-L^2 consistency in (S8). Because $c(P) > 0$ and $\hat{\beta} \xrightarrow{p} \beta$ imply $\mathbb{P}(\hat{\iota} = \iota) \to 1$, the empirically oriented estimator shares the same asymptotic distribution as the fixed-direction estimator in direction a.

Proof of Corollary S3 and the pointwise gap interval. For the curve, combine the fixed-budget limit in Corollary S2 with standard-error consistency in Proposition S3 and apply Slutsky's theorem. For the gap, subtract the margin expansion in (S8) from the curve expansion; the difference has influence (S9). The two fitted scores converge in empirical L^2 by Proposition S3 and (S8), so the second-moment law and Cauchy–Schwarz give $\hat{\sigma}_g^2(p) \xrightarrow{p} \sigma_g^2(p)$. The i.i.d. central limit theorem and Slutsky's theorem give the stated gap coverage. □

S3.3 Curve geometry and functional limit

Fix $\mathscr{P} = [\underline{p}, \overline{p}] \subset (0,1)$ under Assumption S3, and write $\mathbb{G}_N\phi = \sqrt{N}(P_N - P)\phi$ and $\|h\|_{N,2} = (P_Nh^2)^{1/2}$. Probabilities involving multipliers are joint over data and multipliers. Where measurability is unresolved, $\mathbb{P}^*$ and $\mathbb{E}^*$ denote outer probability and outer expectation, respectively.

At a regular contour point, let $\xi(\lambda, p)$ be the upper-p cutoff, $D_{\lambda,p} := \{W(\lambda) \geq \xi(\lambda, p)\}$ its active region and $B(\lambda, p) = \mathbb{E}[Z(a)\mathbf{1}\{D_{\lambda,p}^c\}]$ the retained leverage. Then $\partial_p F = \xi$ and $\partial_\lambda F = -B$ — the cutoff converts criterion uncertainty into share uncertainty, while retained leverage converts it into movement uncertainty (Theorem S1 and corol-

lary S2). Write $\xi_a(p) = \xi(\delta(p), p)$ for the root-score cutoff and $B_a(p) = B(\delta(p), p)$ for retained directional leverage.

Proof of Lemma S1. Recall $f(\lambda; p) = \mathrm{T}_p(W(\lambda)) - \lambda m_Z$.

The cutoff map is continuous on $\mathscr{U}_0$: as $(\lambda', p') \to (\lambda, p)$, the scores converge in L^1, and the strict count gaps at $\xi(\lambda, p) \pm \eta$ trap the moving cutoff for each sufficiently small $\eta > 0$. Cutoff atomlessness and dominated convergence then give continuity of retained leverage and the selected moments.

Derivative in p. For fixed λ, $p \mapsto \int_{1-p}^1 F^{-1}_{W(\lambda)}(u)\, du$ has a moving lower limit; since $F^{-1}_{W(\lambda)}$ is continuous at $1 - p$ for $(\lambda, p) \in \mathscr{U}_0$ (U1), Lebesgue differentiation gives $\partial_p f(\lambda; p) = F^{-1}_{W(\lambda)}(1 - p) = \xi(\lambda, p)$, jointly continuous on $\mathscr{U}_0$.

Derivative in λ. By (20) and U1, $\Gamma(\lambda, \xi; p) = \mathbb{E}[(W(\lambda) - \xi)_+ + p\xi - \lambda Z]$ has a unique cutoff minimizer $\xi(\lambda, p)$ on $\mathscr{U}_0$. Danskin's theorem gives the jointly continuous derivative $\partial_\lambda f(\lambda; p) = \mathbb{E}[Z\mathbf{1}\{W(\lambda) \geq \xi(\lambda, p)\}] - m_Z = -B(\lambda, p)$, with $B(\delta(p), p) = B_a(p) \geq B_{\min} > 0$ by U2.

Implicit function theorem. Since $f(\delta(p); p) = 0$ for each p (Proposition 5) and f is C^1 on $\mathscr{U}_0$ with $\partial_\lambda f \leq -B_{\min}/2 < 0$ on a possibly smaller neighborhood of the root path, the implicit function theorem gives $p \mapsto \delta(p) \in C^1$ with $\partial_p \delta(p) = -\partial_p f / \partial_\lambda f = \xi_a(p)/B_a(p)$, which is (S10). Since $\mathscr{U}_0$ is open, the local implicit solution extends as a C^1 function past each endpoint of $\mathscr{P}$ and agrees with the population root by uniqueness, giving continuity on all of $\mathscr{P}$.

Monotonicity. At the root, normalization and the tail identity give $\mathbb{E}W_a^* = \mathrm{T}_p(W_a^*) = \delta(p)$, so the retained scores $W_a^* \mathbf{1}\{A^{*,c}\}$ have mean zero. Since they lie strictly below $\xi_a(p)$ on a region of mass $1 - p > 0$, the cutoff is positive, and (S10) makes the curve strictly increasing. □

Write $\Delta(p) = \Psi_a(p)/B_a(p)$ and $\mathscr{G}_\Psi = \{\Delta(p) : p \in \mathscr{P}\}$. By Lemma S1 and assumption S3, the population coefficient paths are continuous on $\mathscr{P}$ and B_a is bounded away from zero. The compact parameter representation in Lemma S6 therefore shows that $\mathscr{G}_\Psi$ is a pointwise measurable P-Donsker class with envelope $CH \in L^2(P)$, yielding uniform laws for its squares and pairwise products.

Proof of Theorem S2. The uniform expansion proved in Section S3.2 writes the process as $\mathbb{G}_N \Delta(p) + o_p(1)$ uniformly. The Donsker property of $\mathscr{G}_\Psi$ gives the tight Gaussian limit with covariance (S11). The score Lipschitz bound and compact continuous

parameter path make $p \mapsto \Delta(p)$ uniformly continuous in $L^2(P)$, so the limit has almost surely uniformly continuous paths in p. □

S3.4 Multiplier validity and empirical curve properties

Use the curve-only specialization of the paired process in Section S2.4.1:

$$\hat{\mathbb{G}}_N^*(p) := \hat{\mathbb{G}}_{g,N}^*(p)\Big|_{\hat{\varphi}_c \equiv 0}. \tag{S20}$$

Let $\mathbb{G}_N^\circ$ denote the multiplier process built with the population scores $\Psi_{a,i}(p)$ and slope $B_a(p)$, and write $\hat{\Delta}(p) = \hat{\Psi}_a(p)/\hat{B}_a(p)$ for the fitted score in $\hat{\mathscr{G}}_{\Psi,N}$ of Lemma S6. For i.i.d. standard-normal multipliers independent of the data, the conditional multiplier central limit theorem (van der Vaart and Wellner, 1996, Theorem 2.9.6), applied to $\mathscr{G}_\Psi$, gives

$$\mathbb{G}_N^\circ \rightsquigarrow \mathbb{G} \qquad \text{in } \ell^\infty(\mathscr{P}) \text{ conditionally on the data, in probability.} \tag{S21}$$

Lemma S7 (Uniform plug-ins and fitted multipliers)**.** *Under Assumption S3,*

(i) The uniform plug-in and compact-localization conclusions of Proposition S3 hold on $\mathscr{P}$.

(ii) The fitted multiplier difference is negligible in supremum norm conditionally on the data, in probability, so (S21) *holds for $\hat{\mathbb{G}}_N^*$.*

Proof. The plug-in conclusions follow from Proposition S3 with $\mathscr{I} = \mathscr{P}$.

Multiplier replacement. Given the data, the difference process $\hat{\mathbb{G}}_N^* - \mathbb{G}_N^\circ$ is the Gaussian multiplier indexed by the centered differences $\eta(p) - P_N\eta(p)$, where $\eta(p) = \hat{\Delta}(p) - \Delta(p)$. Adjoin the zero score to anchor the process. Empirical centering is a contraction, so the resulting class has diameter at most $2D_N$, where $D_N = \sup_p \|\eta(p)\|_{N,2} = o_p(1)$. Taking differences of $\varepsilon/4$-nets for $\hat{\mathscr{G}}_{\Psi,N}$ and $\mathscr{G}_\Psi$, then centering, gives a log covering bound $\log 2 + 2\psi(\varepsilon/4)$ by Lemma S6, allowing internal centers and the zero score. On the common covering event intersected with $\{2D_N \le 1\}$, Dudley's bound for centered Gaussian processes (van der Vaart and Wellner, 1996, Chapter 2) gives a conditional expected absolute supremum bounded by

$$C \int_0^{2D_N} \sqrt{1 + 2\psi(\varepsilon/4)}\, d\varepsilon = o_p(1).$$

The event has probability tending to one, so Markov's inequality gives conditional negligibility in probability and hence unconditional negligibility. Conditional Slutsky with (S21) yields the fitted limit. □

Proof of Theorem S3. Adjoining the mean-zero L^2 function φ_c to $\mathscr{G}_\Psi$ gives joint weak convergence of the curve process and $\mathbb{G}_N\varphi_c$. Consequently, $\sqrt{N}(\hat{g} - g)$ converges in $\ell^\infty(\mathscr{P})$ to the continuous Gaussian process $\mathbb{G}_g(p) = \mathbb{G}(p) - Z_c$, where $(\mathbb{G}, Z_c)$ has the covariance of $(\Delta(p), \varphi_c)$. In particular its covariance is $\mathbb{E}[\varphi_g(p)\varphi_g(q)]$ and its variance path $\sigma_g^2(p)$ is continuous.

The oracle multiplier process based on $\varphi_g(p)$ has the same conditional joint limit by the multiplier theorem for this augmented Donsker class. The fitted curve replacement is uniformly negligible by Lemma S7. The margin replacement is constant across budgets and its conditional variance is at most $P_N(\hat{\varphi}_c - \varphi_c)^2 = o_p(1)$, where P_N is the empirical average. Conditional Chebyshev and Slutsky therefore give $\hat{\mathbb{G}}^*_{g,N} \rightsquigarrow \mathbb{G}_g$ in $\ell^\infty(\mathscr{P})$ conditionally on the data, in probability.

By the plug-in part of Lemma S7 and Assumption S2, $\sup_p P_N(\hat{\varphi}_g(p) - \varphi_g(p))^2 = o_p(1)$. Uniform second-moment laws for $\varphi_g(p)$ follow from the curve's compact score family: its squared envelope is integrable, and the cross-products with φ_c have an integrable envelope by Cauchy–Schwarz. A finite parameter mesh and dominated convergence give their uniform laws. Another application of Cauchy–Schwarz now gives $\sup_p |\hat{\sigma}_g^2(p) - \sigma_g^2(p)| = o_p(1)$. Uniform variance consistency and the positive gap-variance floor then give conditional convergence of $\sup_p \left|\hat{\mathbb{G}}^*_{g,N}(p)\right| / \hat{\sigma}_g(p)$ to $\sup_p \left|\mathbb{G}_g(p)\right| / \sigma_g(p)$, the studentized law of the joint gap-process limit above.

Let F_g denote the CDF of $\sup_{p\in\mathscr{P}} \left|\mathbb{G}_g(p)\right| / \sigma_g(p)$. It is continuous by the Gaussian supremum density theorem (Tsirel'son, 1976), applied to the unit-variance studentized process indexed by budgets and signs. Pólya's theorem gives uniform convergence in probability of the conditional CDFs to F_g, hence $F_g(\hat{q}^g_{1-\alpha}) \xrightarrow{p} 1 - \alpha$. For the observed studentized supremum $M_N^{\sup} = \sup_p \sqrt{N}\, |\hat{g}(p) - g(p)| / \hat{\sigma}_g(p)$, continuous mapping gives $F_g(M_N^{\sup}) \rightsquigarrow \mathrm{Unif}(0,1)$. Thus $F_g(M_N^{\sup}) - F_g(\hat{q}^g_{1-\alpha})$ has a limiting distribution continuous at zero, with probability $1 - \alpha$ below zero. By monotonicity of F_g, the event $M_N^{\sup} \le \hat{q}^g_{1-\alpha}$ agrees with the transformed comparison except possibly when $F_g(M_N^{\sup}) = F_g(\hat{q}^g_{1-\alpha})$, whose probability vanishes. □

S3.5 Scalar–curve influence equivalence

Proof of (S12). At the crossing $p = p^{\dagger}$, we have $c = \delta(p)$, so $\hat{c}$ and $\hat{\delta}(p)$ are both within $O_p(N^{-1/2})$ of $\delta(p)$ by Assumption S2 and corollary S2. The secant bound in Lemma S3 and the retained-leverage and cutoff bounds in Lemma S5(b) therefore give

$$\hat{F}_N(\hat{c}, p) = \{B_a(p) + o_p(1)\}\{\hat{\delta}(p) - \hat{c}\} + O_p(N^{-1}),$$

where the $O_p(N^{-1})$ remainder is bounded by the localized order statistic at $\hat{c}$. Use the exact fitted identity following (24), $\hat{A}_b(p) = \hat{L}\hat{F}_N(\hat{c}, p)$, with $\hat{L} \xrightarrow{p} L$. Combining this identity with (S9) yields the influence score $LB_a(p)\varphi_g(p)$. Comparing the two expansions gives

$$N^{-1/2} \sum_{i=1}^{N} \{\varphi_{A,i}(p) - LB_a(p)\varphi_{g,i}(p)\} = o_p(1).$$

The summand is centered and square integrable, so the i.i.d. CLT forces its variance to be zero and hence establishes equality almost surely. □

S4 Sufficient Conditions and Population Examples

We give sufficient conditions for curve regularity (Assumptions S1 and S3) in continuous or mixed designs, allowing for an atom at zero in the contrast index, and in designs with finite covariate support.

Proposition S4 (Continuous and mixed controls)**.** *Let (x, y) be a random vector in fixed dimension d under i.i.d. sampling, fix $a \neq 0$, and suppose $Q = \mathbb{E}[xx'] \succ 0$, $\mathbb{E}\,\|x\|^4 < \infty$, and $P(x'Q^{-1}a = 0) = 0$. Write $y = m(x) + \sigma_0(x)\varepsilon$ with measurable functions m, σ_0 satisfying*

$$\mathbb{E}[(1 + \|x\|^2)m(x)^2] < \infty, \qquad 0 < \sigma_- \leq \sigma_0(x) \leq \sigma_+ < \infty.$$

Suppose $\mathbb{E}[\varepsilon \mid x] = 0$ and the conditional law of ε given x has a jointly measurable density $h_x(t)$ that is strictly positive and continuous in t for almost every x. For deterministic constants $M_h, V < \infty$, assume that for almost every x

$$\sup_{t\in\mathbb{R}}(1+t^2)h_x(t) \leq M_h, \qquad t^2 h_x(t) \longrightarrow 0 \quad (|t| \longrightarrow \infty, \text{ pointwise in } x), \qquad \mathbb{E}[\varepsilon^2 \mid x] \leq V.$$

With $\beta = Q^{-1}\mathbb{E}[xm(x)]$, the standing assumptions Assumptions 2, 3 and 5 and Assumption S1 hold at every $p \in (0,1)$, and Assumption S3 holds on every compact interval $\mathscr{P} \subset (0,1)$. Consequently, both the pointwise intervals and simultaneous multiplier bands apply.

Proposition S5 (Zero-index atoms with verified local cutoff regularity)**.** *Maintain the conditions of Proposition S4 except for $P(x'Q^{-1}a = 0) = 0$. For a compact budget interval $\mathscr{P} = [\underline{p}, \overline{p}] \subset (0,1)$, set $\rho_0 = P(x'Q^{-1}a = 0)$ and assume $\overline{p} < 1 - \rho_0$. Then at each $p \in \mathscr{P}$ a region of mass p attains the curve value at a unique positive cutoff. The standing conditions Assumptions 2 and 3 and regularity conditions Assumptions S1 and S3 hold on $\mathscr{P}$, so the pointwise intervals and simultaneous multiplier bands apply without requiring Assumption 5. For an estimated margin, the gap band still uses the conditions of Theorem S3.*

Proof of Propositions S4 and S5. Write $r_0 = Q^{-1}a$, $\mu = a'r_0$, $t_a = r_0'x$, $Y = t_a u$, and $Z = t_a^2/\mu$. The bounded positive scales, fourth design moment, and exact mixed moment $\mathbb{E}[(1 + \|x\|^2)m(x)^2] < \infty$ imply

$$\beta = Q^{-1}\mathbb{E}[xm(x)], \quad \mathbb{E}[xu] = 0, \quad 0 < \mathbb{E}u^2 < \infty, \quad \mathbb{E}[\|x\|^2 u^2] < \infty.$$

Thus $\mathbb{E}Y = 0$, $\mathbb{E}Z = 1$, $\mathbb{E}(Y^2 + Z^2) < \infty$, and $H = 1 + \|x\|\,|u| + \|x\|^2 \in L^2$, establishing the standing moments and C1.

Leverage and roots before cutoff smoothness. Fix $\mathscr{P} = [\underline{p}, \overline{p}]$ with $0 < \underline{p} \le \overline{p} < 1 - \rho_0$ and put

$$B_{\min} = \int_0^{1-\overline{p}} F_Z^{-1}(t)\,dt > 0,$$

where positivity follows from $P(Z = 0) = \rho_0 < 1 - \overline{p}$. Bottom-quantile allocation gives $\mathbb{E}[ZR] \ge B_{\min}$ for every retention weight $0 \le R \le 1$ with $\mathbb{E}R = 1 - p$ and $p \le \overline{p}$. This bounds scalar leverage across all rules, not retained-matrix rank. Conditional full support on $\{t_a \neq 0\}$ makes Y nonconstant; since $\mathbb{E}Y = 0$, Lemma 1 gives $\mathrm{T}_p(Y) > 0$ for every $p \in (0,1)$. Hence Assumption 3 and Proposition 5 yield a unique positive root for fractional exclusions, $\delta(p) \le \mathbb{E}|Y|/B_{\min}$. Quantile integration is continuous in p for integrable scores, and $|\mathrm{T}_p(W_\lambda) - \mathrm{T}_p(W_{\lambda'})| \le |\lambda - \lambda'|\mathbb{E}Z$. Joint continuity and strict descent with slope at most $-B_{\min}$ therefore give continuity of $p \mapsto \delta(p)$ without cutoff regularity. For every finite λ, the conditional law of $W_\lambda = t_a u + \lambda Z$ given x has a strictly positive continuous density on $\mathbb{R}$ when $t_a(x) \neq 0$ and is a point mass

at zero when $t_a(x) = 0$. When $\rho_0 = 0$ its CDF is continuous and strictly increasing everywhere, proving Assumption 5 for every $\lambda \geq 0$. In either case, at the root,

$$\mathbb{E}[W_\delta] = \delta = \mathrm{T}_p(W_\delta), \qquad \int_0^{1-p} F_{W_\delta}^{-1}(t)\, dt = 0.$$

Because W_δ puts strictly positive mass on $(-\infty, 0)$, a nonpositive $(1-p)$-quantile would make this integral strictly negative. Thus $\xi(p) = \xi_a(p) > 0$, where the CDF is continuous and strictly increasing. The cutoff is unique and the upper-tail region has exact mass p, identifying $\delta(p) = \delta(p)$.

Joint nuisance regularity through zero indices. Let $\theta = (\beta^\circ, r)$, $\mu_r = a'r$, $s = r'x$, and $g = m(x) - x'\beta^\circ + \lambda s/\mu_r$, so

$$W_{\theta,\lambda} = s(y - x'\beta^\circ) + \lambda s^2/\mu_r, \qquad t = (w/s - g)/\sigma_0(x) \quad (s \neq 0).$$

Work on compact parameter sets with $\mu_r \geq \mu/2$, bounded λ, and $w \in [\eta, M]$, for $\eta > 0$. Holding x fixed, the conditional score CDF is $F_x(t)$ for $s > 0$, $1 - F_x(t)$ for $s < 0$, and 1 for $s = 0$. For $s \neq 0$ its derivatives are

$$\begin{aligned} \partial_w F &= h_x(t)/(|s|\sigma_0), & \nabla_{\beta^\circ} F &= \operatorname{sign}(s) h_x(t) x/\sigma_0, \\ \partial_\lambda F &= -|s| h_x(t)/(\mu_r \sigma_0), & \nabla_r F &= \frac{\operatorname{sign}(s) h_x(t)}{\sigma_0}\left[\left(-\frac{w}{s^2} - \frac{\lambda}{\mu_r}\right)x + \frac{\lambda s}{\mu_r^2} a\right]. \end{aligned}$$

They extend continuously by zero through $s = 0$, using $t^2 h_x(t) \to 0$ pointwise in x. Actual differentiability at $s = 0$ follows because, for fixed x and locally small s, the conditional CDF differs from 1 by at most $P(|\varepsilon| > C_x/|s| \mid x) = O(s^2) = o(|s|)$. No density derivative is assumed. The identity $w/s = g + \sigma_0 t$ gives

$$\frac{h_x(t)}{|s|\sigma_0} \leq C(1 + |m(x)| + \|x\|)/\eta, \qquad \frac{w}{s^2} h_x(t) \leq C(1 + m(x)^2 + \|x\|^2)/\eta.$$

All first CDF derivatives therefore admit the integrable envelope $C\{1 + |m(x)| + \|x\|^3 + \|x\|\, m(x)^2\}$, even for unbounded controls. The envelope dominates conditional CDF difference quotients along parameter line segments, including through $s = 0$, so dominated differentiation makes the unconditional CDF jointly C^1. Its density $\mathbb{E}[h_x(t)/(|s|\sigma_0)]$, with the integrand extended by zero at $s = 0$, is finite, continuous, and positive on positive-score domains: $\mu_r > 0$ and $Q \succ 0$ imply $P(s \neq 0) > 0$. No

global density bound at score zero is asserted; nearby nuisance directions may have zero-index atoms.

The quantile implicit function theorem gives locally C^1 nuisance quantiles about each positive root cutoff. Root continuity implies cutoff continuity, hence compactness gives $0 < \min_p \xi(p) \le \max_p \xi(p) < \infty$. A sufficiently small common compact neighborhood of the nuisance parameters, root path, and cutoff path lies strictly in the positive-score region, with density bounded above and away from zero. Quantile bracketing and uniqueness join its local quantile maps, proving C2 and U1.

Leverage, differentiability and empirical expansions. Uniform L^1 continuity of $Z_\theta = (r'x)^2/(a'r)$ preserves retained leverage at least $B_{\min}/2$ on a small nuisance neighborhood, proving C3 and the slope clause of U2. The moments and cutoff regularity already established give C1–C3; Lemma S2 then yields direction differentiability, with the gradient in (S3).

Positive variances and feasible inference. Put $\alpha(x) = m_a(p)'Q^{-1}x$ and $B = B_a(p)$. For a square-integrable score $\Psi_a(p)/B - \gamma(x)u$, the slopes in u on the retained and active regions are $-\alpha(x)/B - \gamma(x)$ and $(t_a - \alpha(x))/B - \gamma(x)$ by (S4). Wherever $t_a \neq 0$, the slopes differ by t_a/B and conditional full support gives the nonzero-slope branch positive probability over an interval. The conditional variance is positive there, so the law of total variance gives positive unconditional variance. This covers both the curve alone ($\gamma = 0$) and the sign-margin gap ($\gamma = t_a$). Continuous root and cutoff paths, cutoff atomlessness, and the common L^2 envelope give continuous selected moments and influence variances, including for the sign-margin gap. Compactness of $\mathscr{P}$ then yields positive variance floors, proving C4 and the variance clause of U2. The fitted expansion and multiplier limit follow from Theorem S2 and Lemma S7. □

Corollary S5 (Paired sign-margin inference)**.** *Under Proposition S4 for a fixed contrast $a_0 \neq 0$ and floor b, suppose $c = |a_0'\beta - b| > 0$ and orient $a = \iota a_0$, where $\iota = \operatorname{sign}(a_0'\beta - b)$. Then (S8) holds with $\varphi_c = (r_a'x)u$ and its fitted counterpart, and*

$$\inf_{p\in\mathscr{P}} \operatorname{Var}\left\{\frac{\Psi_a(p)}{B_a(p)} - (r_a'x)u\right\} > 0.$$

Thus Theorem S3 applies. Validity is pointwise in the population law, not uniform as c approaches zero.

Proof of Corollary S5. The sign consistency and OLS expansion in Section S3.2 give (S8) with $\varphi_c = (r_a'x)u$. On the event of correct orientation, the fitted-minus-population margin score is bounded by $o_p(1)H$, giving the required empirical-L^2 consistency. The conditional-branch argument in Proposition S4 with $\gamma(x) = r_a'x$ gives the gap-variance floor, so Theorem S3 applies. □

For example, when $x \sim \mathrm{N}(0, I_d)$ and $u \sim \mathrm{N}(0,1)$ are independent, $y = x'\beta + u$, and $a = e_1$, Proposition S4 applies and the design gradient vanishes by Corollary S4. The active set depends only on $T = x_1$ and u, so the sole correction to the oracle numerator in (S4) is $-\mathbb{E}[T^2\mathbf{1}\{A_p^*\}]Tu$.

S4.1 Finite covariate support

Proposition S6 (Finite-covariate primitive verification)**.** *Let x take fixed values $x_1, \dots, x_J \in \mathbb{R}^d$ with probabilities $\pi_j > 0$, where J and d are fixed and $Q = \sum_{j=1}^J \pi_j x_j x_j' \succ 0$. For each $j \in \{1, \dots, J\}$, suppose that $y \mid x = x_j$ has a strictly positive continuous density h_j on $\mathbb{R}$ and a finite second moment. Fix $a \neq 0$ and suppose $t_{a,j} = x_j'Q^{-1}a \neq 0$ for every j. Under i.i.d. sampling of (x, y), with β the linear projection coefficient of (13), Assumption S1 holds at every $p \in (0,1)$, and Assumption S3 holds on every compact interval $\mathscr{P} \subset (0,1)$. In particular, with $\mu = a'Q^{-1}a$,*

$$B_a(p) \geq (1-p) \min_{1 \leq j \leq J} \frac{t_{a,j}^2}{\mu} > 0.$$

Proof of Proposition S6. Write $c_j = \mathbb{E}[y \mid x = x_j]$, $\mu = a'Q^{-1}a$, and $r_a = Q^{-1}a$. The projection coefficient is $\beta = Q^{-1}\sum_{j=1}^J \pi_j x_j c_j$, so $\mathbb{E}[xu] = 0$; finite support and finite conditional second moments give C1 and the standing model moments. Put $z_j = t_{a,j}^2/\mu > 0$. Every retained region has leverage at least $(1-p)\min_{1\leq j\leq J} z_j$, so Assumption 3 holds. The conditional densities h_j imply that $Y(a) = t_{a,j}u$ (conditional on $x = x_j$) has a continuous, nondegenerate full-support distribution and unconditional mean zero, so Lemma 1 makes its upper-p tail integral positive. Thus Proposition 5 gives a positive finite root.

Joint cutoff regularity. For $\theta = (\beta^\circ, r)$ near (β, r_a), let $s_j = r'x_j$, $\mu_r = a'r$, and

$$W_{\theta,\lambda} = (r'x)(y - x'\beta^\circ) + \lambda\frac{(r'x)^2}{\mu_r}, \qquad \gamma_j = -s_j x_j'\beta^\circ + \lambda s_j^2/\mu_r.$$

Choose a compact neighborhood where $\mu_r \geq \mu/2$ and each s_j is bounded away from zero with the sign of $t_{a,j}$. In cell j, the score is $s_j y + \gamma_j$, with mixture density

$$f_{\theta,\lambda}(w) = \sum_{j=1}^{J} \frac{\pi_j}{|s_j|} h_j\left(\frac{w-\gamma_j}{s_j}\right). \tag{S22}$$

This density is positive and jointly continuous in (θ, λ, w), and on compact parameter and cutoff sets it is bounded above and away from zero. The cellwise CDFs are $F_j((w-\gamma_j)/s_j)$ for $s_j > 0$ and $1 - F_j((w-\gamma_j)/s_j)$ for $s_j < 0$; since $F_j' = h_j$ is continuous, the mixture CDF is C^1 and strictly increasing. Positivity and continuity hold at the population nuisance value for every $\lambda \geq 0$, giving Assumption 5; only the uniform density bounds require compactness.

Fix $\mathscr{P} = [\underline{p}, \overline{p}] \subset (0,1)$. The retained-leverage lower bound is $B_{\min} = (1-\overline{p}) \min_{1\leq j\leq J} z_j > 0$, so the root is bounded by $\mathbb{E}|Y(a)|/B_{\min}$. Joint continuity of the tail criterion on bounded root sets and this uniform descent give continuity of the root path, and strict quantile monotonicity gives continuity of its cutoff path. On a common band around that compact path, the density in (S22) has uniform positive lower and finite upper bounds. Quantile bracketing and the implicit function theorem therefore yield a locally C^1 cutoff map, verifying C2, U1, and the slope clauses of C3 and U2.

Process regularity and variances. The conditional-branch argument of Proposition S4 applies with full support and $t_{a,j} \neq 0$ in every cell, giving positive fitted-score variance. The common envelope $C(1+|y|) \in L^2(P)$, continuous root and cutoff paths, and cutoff atomlessness give continuous variances and positive compact floors, proving C4 and U2. The fitted expansion then follows from Theorem S2. For a fixed floor with positive oriented margin, the same conditional-branch argument with $\gamma(x) = t_a$ gives the positive gap-variance floor needed to apply Theorem S3 with the margin expansion (S8). □

S4.2 Matched signal-to-noise populations

The Gaussian and logistic populations in Table S1 share the covariate law, conditional means and conditional variances, matching ϑ and V below within 10^{-12}. Their different sensitivity curves show what standardized signal-to-noise leaves undetermined (Section 5).

TABLE S1: Different tipping fractions at the same coefficient and asymptotic standard error

Design	ϑ	$\sqrt{V}$	$\delta(0.02)$	$\delta(0.10)$	$\delta(0.20)$	$p^\dagger$
Gaussian errors	0.2036	2.1461	0.1211	0.4523	0.7987	0.0375
Logistic errors	0.2036	2.1461	0.1293	0.4540	0.7839	0.0356

Notes: $\vartheta = a_1\beta$ and $V = \mathbb{E}[(a_1 Q^{-1} xu)^2]$. The sign crossing (zero floor $b = 0$, positive movement margin $c = \vartheta$) is $p^\dagger = \inf\{p : \delta(p) \geq \vartheta\}$. All rows are deliberately calibrated to $\vartheta/\sqrt{V/1000} = 3$ (absolute tolerance 10^{-12}); 1000 is a reference size, not a realized sample or t-statistic. The two populations also match in ϑ and V separately within 10^{-12}. The curve values and crossings are numerical solutions for the specified populations, not estimates from sampled data.

There are four strata s and a binary treatment T, with $x = (1, T, \mathbf{1}\{s = 2\}, \mathbf{1}\{s = 3\}, \mathbf{1}\{s = 4\})'$ and contrast $a = (0, 1, 0, 0, 0)'$. Write $w_s = P(s)$ and $q_s = P(T = 1 \mid s)$, so the eight cell probabilities are $w_s(1 - q_s)$ and $w_s q_s$. Outcomes obey $y = b_s + \kappa_{\text{eff}} g_s T + \sigma_s e$, where e is independent of (s, T), has mean zero and variance one, and $b = (0.40, -0.25, 0.70, -0.35)$. Set $w = (0.30, 0.30, 0.20, 0.20)$, $q = (0.5, 0.3, 0.7, 0.4)$, $g = (0.6, -0.1, 0.9, -0.4)$ and $\sigma = (1, 1, 1, 1)$. The Gaussian specification uses standard normal e; the logistic specification scales a standard logistic error by $\sqrt{3}/\pi$ to preserve unit variance. In each specification, β is the population linear projection coefficient, $u = y - x'\beta$, and $\vartheta = a'\beta$. The asymptotic variance of $\sqrt{N}(\hat{\vartheta} - \vartheta)$ is $V = \mathbb{E}[(a'Q^{-1}xu)^2]$. The common effect scale $\kappa_{\text{eff}} \approx 0.81011$ sets the population signal-to-noise ratio $\vartheta/\sqrt{V/1000}$ to 3 at $N = 1{,}000$.

All eight cell probabilities are positive, and the covariate support spans $\mathbb{R}^5$. Writing $L_T = \sum_{s=1}^{4} w_s q_s (1 - q_s) > 0$, the cell index is $x'Q^{-1}a = (T - q_s)/L_T \neq 0$. Gaussian and logistic conditional densities are continuous and strictly positive on $\mathbb{R}$, with finite second moments; hence both laws satisfy Proposition S6, including its positive retained-denominator guarantee. The table entries solve the population modified-score cutoff and ratio equations of Section B, up to the reported numerical optimization tolerance.

With logistic rather than Gaussian errors, the worst-case movement is larger at budgets 0.02 and 0.10 but smaller at 0.20, and the sign crossing occurs earlier.

S5 Calibration Designs and Evidence

Section E.1 reports the primary scalar experiment, how replications without an interval enter coverage, and the resulting coverage bounds. Here we specify the law

constructions and numerical guards, and then present the designs and results for the additional scalar and estimated-margin band experiments.

Mindspark-based populations. For each subject separately, let (x_i, y_i) be the complete-score analysis rows of Section D.1, $\hat{\beta}$ the reference OLS coefficient, $u_i = y_i - x_i'\hat{\beta}$, and $s^2 = N^{-1}\sum_i u_i^2$. Draw design rows independently and uniformly from the empirical design distribution. Given a draw of row x_i, the normal law generates $y = x_i'\hat{\beta} + s\varepsilon$, with independent standard-normal ε. The heteroskedastic law replaces s with $s_i = K\exp\{\gamma(c_i - \bar{c})/2\}$, where c_i is baseline achievement,

$$\gamma = \mathrm{clip}_{[-0.75,0.75]}\left\{\frac{\mathrm{Cov}_N(c_i, \log(u_i^2 + 0.05s^2))}{\mathrm{Var}_N(c_i)}\right\}, \qquad N^{-1}\sum_i s_i^2 = s^2,$$

and $K > 0$ enforces the second equality. The heavy-tail law uses $s\varepsilon/\sqrt{3}$ for independent Student-t_3 errors, matching the residual variance. The rounded law rounds the normal-law outcome to increments of 0.1 and clips it to endpoints obtained by rounding outward the minimum and maximum of $x_i'\hat{\beta} \pm 9s$. To compute the population target, we account for the probability masses at the clipping endpoints and recompute the population linear projection after rounding and clipping. The empirical-residual law draws a centered residual independently of the design row and adds it to $x_i'\hat{\beta}$.

STAR-based populations. Use the 729×32 design of Section D.2, with the contrast on the combined-offer indicator. Let $m_i = x_i'\hat{\beta}$, let s_g be residual root mean square within each of the four offer arms, and let $g(i)$ denote the arm of observation i. With independent standard-normal ε or Student-t_5 noise, the six laws are:

(i) $y_i + 0.25\varepsilon$;
(ii) $y_i + \varepsilon$;
(iii) linear normal: $m_i + s_{g(i)}\varepsilon$;
(iv) linear heavy-tail: $m_i + s_{g(i)}\sqrt{3/5}\,t_5$;
(v) nonlinear heteroskedastic: $y_i + s_{g(i)}\varepsilon$;
(vi) tilted nonlinear heteroskedastic: $y_i + s_{g(i)}\varepsilon$, with row sampling weights halved for rows in at least one control category observed at most five times, then normalized.

Laws (i)–(v) draw design rows independently and uniformly; law (vi) draws them independently with the tilted weights. Simulations operate in an estimable design basis while preserving the original contrast.

S5.1 Numerical tolerances and checks

For the targets in Section E.1, Mindspark equation residuals are at most 10^{-9}, continuous-law curve values at three budgets agree with an independent scalar-root solver within 10^{-8}, and finite-law iteration uses a 2×10^{-13} step tolerance. STAR targets require equation residuals at most 10^{-7} and positive minimum population retained leverage, and the synthetic crossings agree with an independent scalar-root solver within 10^{-9}.

The sample support calculation requires $\hat{L} > 10^{-12}$ and $\min_{|S|=k_{\max}} \sum_{i \notin S} \hat{v}_i^2 > 10^{-10}$, where $k_{\max}$ is the largest padded count. The directional fit requires positive retained leverage and a numerical gap estimate at most 10^{-8} at each fitted node.

S5.2 Additional scalar calibration

In the primary four-cell experiment, coverage with t_3 errors at $N = 200$ is 94.34%, with simultaneous bounds [93.63%, 95.00%]. A separate experiment with 100,000 replications yields 94.679% with bounds [94.538%, 94.817%], exceeding the prespecified 94% practical floor for that law without changing the simultaneous family bounds.

The following experiments use the same padded-cell rule, target crossings clipped to the reporting range, and count replications without an interval as noncoverage. For a target crossing $p^\dagger$, set $\mu = \phi(\Phi^{-1}(1 - p^\dagger))/(1 - p^\dagger)$.

Crossings at and beyond reporting boundaries. Two six-law families use $N = 200$, 12,000 replications per law and reporting range $[0.103, 0.347]$. Their true crossings are $0.200, 0.2025, 0.103, 0.347, 0.080, 0.380$, spanning count nodes, cell interiors, reporting boundaries, and shares outside the reporting range. The fixed-floor family uses $y = \mu + \varepsilon$ with an intercept and floor zero. The estimated-half-floor family targets the intercept in the regression of $y = 2\mu + X + \varepsilon$ on $(1, X)$, for independent standard-normal X, ε, with fitted floor $\hat{b} = \hat{\beta}_0/2$. The population floor is μ and the residualized intercept is 1, so both families have the same population support score $\mu + \varepsilon$ and use the same exact calibration. In both families, all within-family lower coverage bounds exceed 0.94.

Weak orientation. The latter construction at $N = 729$ uses reporting range $[0.005, 0.10]$, crossings $0.004, 0.005, 0.006, 0.050, 0.100, 0.101$ and 12,000 replications per law. At crossings 0.4%, 0.5% and 0.6%, unconditional coverage is 72.31%, 76.18% and 80.73%, respectively. The fitted contrast fails the prespecified orientation check and yields no interval in 26.12%, 21.75% and 17.45% of replications, respectively. At the same sample size, smaller target crossings have smaller positive reference margins.

S5.3 Estimated-margin band calibration

The band experiment uses the four-cell design of Section E.1 with a fixed floor of 1, independent normal or unstandardized Student-t_3 errors, 1,600 replications per cell, and 999 common multipliers. In each replication, the procedure estimates the orientation and margin and evaluates coverage against the fixed population orientation, counting an undefined orientation as noncoverage. Every replication had a defined orientation, and the estimated sign flipped in 26 of the 1,600 replications in the t_3, $N = 200$ cell. Conditional-tail integrals supply population curve values at both ends of each count cell on $[0.02, 0.20]$.

Estimated simultaneous coverage at the starts of the count cells exceeds coverage over the full reporting range by 0.13–0.88 percentage points. Table S2 reports full-interval coverage for the band constructed with the multiplier critical value and no additional width adjustment.

TABLE S2: Simulated coverage of simultaneous bands for worst-case movement minus the conclusion margin

Law	N	Gap-band coverage
Normal	200	0.932
	800	0.936
	3 200	0.952
t_3	200	0.927
	800	0.943
	3 200	0.946

Notes. A nominal 95% band counts as covering if it contains the gap at every exclusion share from 2% to 20%. Each law and sample size uses 1,600 replications with 999 common multipliers. Coverage is a fraction of all attempts. Every attempt fits the regression and produces an estimated orientation.

Dependence diagnostic. Shared-stratum shocks caused undercoverage in a separate Mindspark band-inversion stress that allocated 10% of marginal error variance to shared shocks. This serves as evidence against treating student dependence as harmless, rather than a calibration of the current scalar interval under dependence; the scalar and band results above maintain i.i.d. sampling.

S6 Additional Empirical Analyses

Section D.1 specifies the usable-score sample, scoring map, and observed-case population model. The 31 half-effect students — 15 offered and 16 controls — occupy baseline quartiles in counts 9, 5, 7 and 10. Their arm-specific outcome residuals average +1.360 (offered) and −1.647 (control), and the retained fraction of identifying variation is 0.9424. Table S3 compares these exclusions with pooled baseline tails, adverse residuals, the fitted influence ranking ("frozen influence" in the table), and 2,000 random deletions.

TABLE S3: Adjusted Mindspark math gains after equal-sized exclusions

Deletion rule	Fixed-adjustment coefficient	Full-regression refit
Top half-effect support scores	0.18447	0.18426
Lowest baseline	0.40489	0.41197
Highest baseline	0.37605	0.37627
Adverse residual	0.18458	0.18426
Frozen influence	0.18447	0.18426
Random (median)	0.36930	0.36935

Notes. Each rule removes 31 of 535 students with usable math scores. Coefficients are in baseline standard deviations, and half the original gain is 0.18467. The baseline exclusions are pooled, not arm-balanced. Adverse residuals rank the arm-signed baseline residual; frozen influence ranks the unmodified influence score. Random rows summarize 2,000 equal-sized deletions. Refits are attained sample diagnostics on the chosen complements, not unrestricted optima or post-selection confidence statements.

Arm-restricted sample audits.. Arm-restricted audits hold the reference adjustment and regression specification fixed. For exact group counts $(k_g)_g$, sort the fitted support scores within each group from largest to smallest, $\hat{q}_{g,[1]} \geq \hat{q}_{g,[2]} \geq \cdots$. If every eligible deletion set leaves positive retained identifying variation, some deletion with those

counts overturns the conclusion exactly when

$$\sum_{g} \sum_{j=1}^{k_g} \hat{q}_{g,[j]} \geq \sum_{i} \hat{q}_i.$$

Increasing one group's exact count while holding the others fixed can force removal of a negative score, reducing the deleted support. If k_g is a cap instead, delete only positive scores among that group's top k_g; raising the cap cannot reduce the maximum deleted support. The reported arm-only minima are the first successful exact counts, with all other counts fixed at zero.

Mindspark's arm-only half-effect minima are 46/264 offered students (17.42% of that arm) or 41/271 controls (15.13%), compared with the pooled 31/535. STAR's focal-only half-effect minimum is 8/62 (12.9%), and deleting only from the other arms requires 31/667 (4.65%). The focal-only minimum matches the sample minimum with all arms eligible because the attaining set in Section 4.2 consists of eight combined-offer women.

References

Tsirel'son, V. S. (1976). "The Density of the Distribution of the Maximum of a Gaussian Process". In: *Theory of Probability & Its Applications* 20.4, pp. 847–856. DOI: 10.1137/1120092.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer. DOI: 10.1007/978-1-4757-2545-2.