Tipping Fractions for Regression Conclusions

Lucas D. Konrad
WU Vienna
lucas.konrad@wu.ac.at

Nikolas Kuschnig
Monash University
nikolas.kuschnig@monash.edu

Preliminary Draft
November 24, 2026

Regression estimates can be precisely estimated, yet rest on a small share of the data. We study the *population tipping fraction*, the smallest share whose exclusion overturns a stated conclusion with reference adjustments held fixed. Ranking observations by their support for the conclusion yields the minimum sample deletion count, and interpolation estimates the population share. We derive its limiting distribution and construct a confidence interval that accounts for using the same sample to fit the adjustments, estimate a relative threshold, and identify the most supportive observations. In the Mindspark experiment, excluding 31 of 535 students cuts the estimated math gain by at least half. The nominal 95% interval for the tipping fraction is 3.36–8.60%, and falls below a 13.6% attrition benchmark.

1. Introduction

A regression estimate can be precise without being broadly supported by the data. In the Mindspark education experiment, which offered students an after-school program with personalized software and instructor support, the estimated effect on math achievement is positive and statistically significant (Muralidharan, Singh, and Ganimian, 2019). Excluding 31 of the 535 assessed students (5.79%) cuts the estimated gain by at least half, holding the full-sample baseline adjustments fixed. What share of the population would have to be lost to halve the gain? How much could erase it, and how precisely can we infer those shares from the data?

We study the *population tipping fraction*, the smallest share of a reference population whose selective exclusion overturns a specified regression conclusion under the reference adjustment.[1] The conclusion may be that a coefficient is positive, exceeds a policy threshold, or retains more than a specified fraction of its reference magnitude. We measure how robust this conclusion is to shifts in who is represented by the analysis, e.g., due to selective attrition or nonresponse. If the share at risk falls below the tipping fraction, it is too small to overturn the conclusion, even if the loss is as adverse as possible.

We use a *common reference adjustment* to evaluate changes in representation against a fixed benchmark. To see what stays fixed, consider plotting endline achievement against treatment offer in Mindspark, adjusting both variables for baseline achievement in the full sample. Remove some students, leaving the remaining points at their original coordinates, and fit a new slope through the origin. This measures how the remaining students support the original study's adjusted comparison. Refitting the full regression would also use those students to re-estimate the baseline adjustments, while holding the coordinates fixed lets us find the smallest deletion with one ranking per conclusion.

[1] Reporting a "third number" alongside the estimate and standard error also motivates Giordano, Meager, and Broderick (2026a,b) for sample deletion and Spini (2026) for robustness to covariate shifts.

1.1. Contributions

Our starting point is an exact link between a sample audit and a population quantity. Each observation has a signed contribution to the conclusion of interest. When net support is positive and some observations oppose the conclusion, removing the largest contributions until they exhaust net support gives the exact minimum deletion count. Interpolating within the last deletion gives an estimator of the population tipping fraction. This characterization draws on the score-over-curvature representation of Konrad and Kuschnig (2026a) and the conditional value-at-risk of Rockafellar and Uryasev (2000).

Our main contribution is inference for the population tipping fraction. The same observations determine the fitted adjustments, any threshold expressed as a fraction of the reference coefficient, and which records are most supportive. Selecting the most supportive observations generally breaks orthogonality, and estimating the adjustment and floor can affect the tipping estimator at first order. We derive a joint influence function and feasible variance estimator that account for these sources of variation, and test whether loss of a given share can exhaust support for the conclusion. Inversion yields two interpretable ranges: below the lower confidence bound, every admissible loss preserves the conclusion; above the upper bound, some admissible loss overturns it.

In Mindspark, 13.6% of assigned students fall outside the math regression. We use that percentage as a benchmark for selective loss among students with usable scores. whether loss of a given share can exhaust support for the conclusion. It exceeds the upper confidence bound for halving the math gain and lies inside the interval for erasing it (see Figure 1). As a result, under the maintained model, some loss of 13.6% could halve the gain in the population, but whether it could erase the gain remains open. Two further sample audits show how support can concentrate in one treatment arm (Angrist, Lang, and Oreopoulos, 2009) and how a loss of identifying variation can substantially change the tipping count relative to a first-order approximation (Nunn and Puga, 2012).

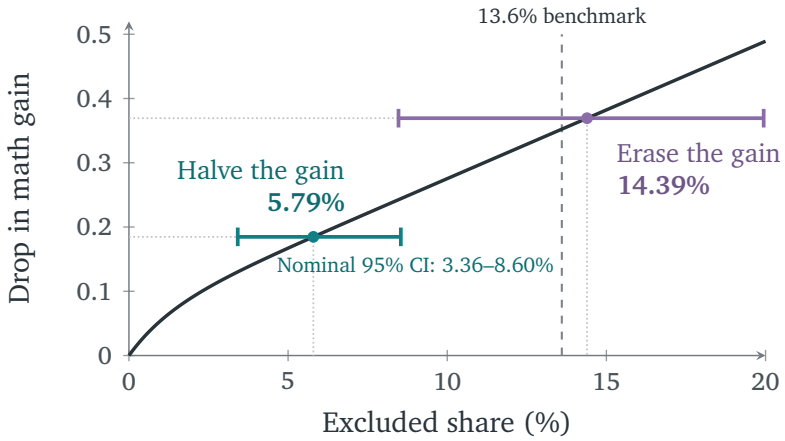

FIGURE 1: Tipping shares for the adjusted Mindspark math gain. Dots mark the minimum sample exclusions that halve the gain (31 students) or bring it to zero or below (77 of 535). The curve joining the two thresholds is schematic. Horizontal segments give nominal 95% intervals for the population tipping fractions within the reporting range, and the dashed line marks the loss benchmark. The sign interval reaches the 20% upper reporting limit.

1.2. Related literature

Our analysis connects to work on subset influence, sensitivity and breakdown inference, and robust statistics. Case-deletion diagnostics ask which observations drive a fitted result (Chatterjee and Hadi, 1986; Cook, 1979), while work on influential subsets examines collective leverage and the limits of single-observation heuristics (Hu et al., 2024; Koh and Liang, 2017; Kuschnig, Zens, and Crespo Cuaresma, 2021), motivating search algorithms and certificates (Moitra and Rohatgi, 2023; Rubinstein and Hopkins, 2024). The Approximate Maximum Influence Perturbation (AMIP) of Giordano, Meager, and Broderick (2026a,b) provides a first-order approximation to the largest adverse movement from deleting a subset and refitting. The exact coefficient change under common adjustment shares AMIP's influence numerator and divides it by the fraction of retained identifying variation, which can alter both the magnitude of the movement and the maximizing subset. Konrad and Kuschnig (2026a) optimize this ratio for fixed residualized inputs, and Konrad

and Kuschnig (2026b) use extreme-value limits to test for excessive maximal influence among fixed-size and vanishing-share sets. Here, we develop population inference for the tipping fraction with non-vanishing loss.

Our reporting goal is shared by work on breakdown frontiers (Masten and Poirier, 2020), policy conclusions under covariate shifts (Spini, 2026), and worst-case effects over covariate-defined subpopulations (Jeong and Namkoong, 2020). Here exclusion can depend on outcomes as well as covariates, placing the retained distributions in the contaminated-sampling neighborhood of Horowitz and Manski (1995). In concurrent work, Palomba (2026) develops sensitivity bounds and breakdown inference for trimmed expectations in that neighborhood. If adjustments and floor were known, our tipping fraction is the first share at which Palomba's lower trimmed-expectation bound reaches zero; the corresponding known-score asymptotics follow from Mafusalov, Shapiro, and Uryasev (2018). We account for fitting both adjustments and any relative floor, and obtain the interval by testing the balance locally at the crossing. Other related sensitivity bounds address missing data (Kline and Santos, 2013; Ober-Reynolds, 2026) and counterfactuals under changes in latent-variable distributions (Christensen and Connault, 2023).

Our focus on adverse exclusions of a nonvanishing population share draws on the robust-statistics tradition of breakdown and maximum-bias curves (Berrendero and Zamar, 2001; Donoho and Huber, 1983; Hampel, 1974; Huber, 1981). Other approaches study the variance and bias of estimators under local misspecification (Armstrong and Kolesár, 2021; Bonhomme and Weidner, 2022; Kitamura, Otsu, and Evdokimov, 2013), local sensitivity to estimating moments (Andrews, Gentzkow, and Shapiro, 2017), distributional derivatives (Ichimura and Newey, 2022), and robust estimation by model-friendly reweighting (Dewaskar et al., 2025).

Roadmap. Section 2 defines the conclusion and its tipping fraction. Section 3 develops the estimator, inference, and calibration. Section 4 presents three applications. Section 5 discusses, and Section 6 concludes. The Appendix gives the proofs and implementation details. The Online Supplement develops

the finite-type and stratified results used for Mindspark, inference for the sensitivity curve, and additional results.

2. Regression Conclusions and Their Support

To measure support for a regression conclusion, we must specify which coefficient to compute after an exclusion, how far it must fall to overturn the conclusion, and which exclusions are allowed.

2.1. Fixed reference adjustments

Consider the coefficient on a regressor w in the linear projection of an outcome y on w and controls z, under a reference population law P with finite second moments and nonsingular design. We write $P(\cdot)$ for probabilities under that law and $\mathbb{P}(\cdot)$ for probabilities over repeated samples. For example, w may be a treatment offer and z a vector containing an intercept and baseline characteristics.

By the Frisch–Waugh–Lovell theorem, this projection coefficient equals the slope in the population added-variable plot. The reference adjustments

$$\pi := \mathbb{E}[zz']^{-1}\mathbb{E}[zw], \qquad \eta := \mathbb{E}[zz']^{-1}\mathbb{E}[zy],$$

define the plot's coordinates: the residualized regressor $v := w - z'\pi$ and the outcome $h := y - z'\eta$. The coefficient $\vartheta = \mathbb{E}[vh]/L$ is the slope through the origin, where $L := \mathbb{E}[v^2]$ is the identifying variation and $u := h - \vartheta v$ is the reference residual. For a retention weight $R \in [0, 1]$, the retained slope through the origin is

$$\vartheta(R) := \arg\min_t \left\{ \mathbb{E}\left[R(h - tv)^2 \right] \right\} = \frac{\mathbb{E}[R\, vh]}{\mathbb{E}[R\, v^2]},$$

defined when $\mathbb{E}[R\, v^2] > 0$.

Exclusion leaves the coordinates of the retained points fixed. At any slope t, the implied fitted value is $tv + z'\eta = tw + z'(\eta - t\pi)$, with control coefficients $\eta - t\pi$. Among all coefficient vectors with the focal coefficient fixed at t, this path minimizes the mean squared change in fitted values from the original regression, averaging over the reference population. (Proposition 3 gives the result for any linear contrast.) The path preserves a common benchmark across exclusions.

2.2. Conclusions and signed support

We audit a *conclusion* of the form $\vartheta > b$ and orient the contrast so that decreases are adverse. Set $b = 0$ to audit a positive coefficient, use a fixed b for a policy threshold, or set $b = \kappa\vartheta$ to ask whether more than a fraction $\kappa \in (0, 1)$ of a positive reference coefficient survives. The margin $c := \vartheta - b > 0$ is the change that would overturn the conclusion, with sample equivalent $\hat{c} = \hat{\vartheta} - \hat{b}$.

Each observation contributes a *signed support score*

$$q_b := v(h - bv), \qquad \mathbb{E}[q_b] = cL > 0.$$

Positive scores support a slope above the floor, and negative scores offset that support. The mean score measure cL is net support for the conclusion. Because $\vartheta(R) - b = \mathbb{E}[Rq_b]/\mathbb{E}[Rv^2]$,

$$\vartheta(R) > b \quad \iff \quad \mathbb{E}[Rq_b] > 0, \tag{1}$$

whenever $\mathbb{E}[Rv^2] > 0$. With exclusion weight $d = 1 - R$, the conclusion fails once deleted support exhausts net support: $\mathbb{E}[dq_b] \geq \mathbb{E}[q_b]$.

For intuition, take a balanced treatment comparison with only an intercept as control and a zero floor $b = 0$. Then $v = \pm 1/2$ and $q_0 = \pm(y - \mathbb{E}[y])/2$. High treated outcomes and low control outcomes support a positive effect, whereas low treated outcomes and high control outcomes oppose it. If these two groups of scores nearly offset, removing just a few large contributions can exhaust net support.

2.3. The population tipping fraction

An *admissible exclusion* of share p is a measurable weight $d = d(y, w, z) \in [0, 1]$ with $\mathbb{E}[d] = p$. Let $\mathscr{D}_p$ denote the class of all such weights. Selection may use the full observed record, including outcomes. Fractional weights can split probability mass at a tied score, and indicator weights correspond to subsets. If the conclusion survives every exclusion in $\mathscr{D}_p$, it also survives narrower rules, such as baseline-only loss or loss confined to one treatment arm.

Definition 1 (Population tipping fraction)**.** *For the conclusion $\vartheta > b$, define*

$$p^\dagger := \inf\left\{ p \in (0,1) : \ \exists d \in \mathscr{D}_p,\ \mathbb{E}\left[(1-d)v^2\right] > 0,\ \vartheta(1-d) \le b \right\},$$

with $\inf \varnothing = +\infty$.

The constraint $\mathbb{E}[(1-d)v^2] > 0$ rules out exclusions that eliminate all identifying variation and leave the retained slope undefined, while $\inf \varnothing = +\infty$ covers cases where no admissible exclusion overturns the conclusion. This is a conclusion-specific breakdown point for selective exclusion, where the floor specifies what counts as breakdown.

In a sample of size N, the sample counterpart fixes the reference adjustments and evaluates the retained slope $\hat{\vartheta}_{-S}$ after deleting an observation subset $S \subset \{1, \dots, N\}$. The *sample tipping count* $k^\dagger$ is the smallest cardinality $|S|$ that preserves positive retained variation and achieves $\hat{\vartheta}_{-S} \le \hat{b}$.

2.4. From sample counts to population shares

The count $k^\dagger$ is exact for the observed sample. A new draw from P changes the fitted adjustment and the observations available for deletion, so $k^\dagger/N$ varies across samples even though $p^\dagger$ stays fixed. Because $p^\dagger$ is a feature of P, it can be compared with budgets for attrition, nonresponse, or selective reporting in that population.

Example 1 (Gaussian location)**.** *With no controls and the intercept as the target coefficient, $v = 1$, and the support score is simply $y - b$. If $y \sim \mathrm{N}(\vartheta, \sigma^2)$ with*

$\vartheta > 0$ and the floor is $b = 0$, deleting the upper tail $y > \vartheta + t\sigma$ brings the retained mean to zero when

$$\frac{\phi(t)}{\Phi(t)} = \frac{\vartheta}{\sigma}, \qquad p^{\dagger} = 1 - \Phi(t),$$

where ϕ and Φ are the standard-normal density and distribution function. At $\vartheta/\sigma = 0.1$, the tipping fraction is about 4.5%. With $N = 1{,}000$, the reference signal-to-noise ratio $\sqrt{N}\,\vartheta/\sigma$ is about 3.2, while removing the most supportive 4.5% of the population exhausts the positive mean. Increasing N sharpens estimates of $p^{\dagger}$ without changing the population target.

3. Characterization, Estimation and Inference

At a given exclusion share, the most support one can remove comes from the upper tail of the score distribution. This observation gives a population characterization, a sample calculation, and a way to test candidate tipping fractions.

3.1. Tipping as exhaustion of net support

For an integrable U and $s \in (0,1)$, let $F_U^{-1}(s) := \inf\{t : F_U(t) \geq s\}$ be its left-continuous quantile function. The upper-tail integral

$$\mathrm{T}_p(U) := \int_{1-p}^{1} F_U^{-1}(s)\, ds \tag{2}$$

sums the largest values occupying population share p. Equivalently, it is the largest expected value that can be removed by excluding a share p:

$$\mathrm{T}_p(U) = \sup_{d \in \mathscr{D}_p} \{\mathbb{E}[dU]\} = \min_{\xi \in \mathbb{R}} \left\{\mathbb{E}[(U-\xi)_+] + p\xi\right\}.$$

The second equality is the upper-tail allocation duality (Rockafellar and Uryasev, 2000), and fractional weights split mass when needed. The *support*

balance

$$A_b(p) := \mathrm{T}_p(q_b) - \mathbb{E}[q_b], \tag{3}$$

is the excess of maximal deleted support over the original net support. When the balance is negative, even the worst exclusion leaves positive net support and the conclusion remains intact. At zero, that support is exhausted.

The upper-tail integral is concave in p, as increasing the exclusion budget adds observations that are progressively less supportive. With positive net support for the conclusion and some negative scores, the support balance has a unique zero in $(0, 1)$ and stays positive between that crossing and 1.

Proposition 1 (Tipping by support exhaustion)**.** *Suppose* $\mathbb{E}[|q_b|] < \infty$*,* $\mathbb{E}[q_b] > 0$ *and* $P(q_b < 0) > 0$*. Then:*

(i) *The tipping fraction of Definition 1 is the first share at which maximal deleted support exhausts net support:*

$$p^\dagger = \inf\{p : A_b(p) \geq 0\},$$

and satisfies $0 < p^\dagger < P(q_b > 0)$*;*

(ii) *for every* $p < p^\dagger$ *and every* $d \in \mathscr{D}_p$*, retained variation is positive and* $\vartheta(1 - d) > b$*;*

(iii) *for every* $p \in [p^\dagger, 1)$*, deleting the upper p-tail of* q_b *(splitting cutoff mass if necessary) leaves positive retained variation and* $\vartheta(1 - d) \leq b$*; at* $p = p^\dagger$*, this exclusion removes only positive support;*

(iv) *the tipping fraction is the value of the minimum-exclusion problem and its scalar dual:*

$$p^\dagger = \min_{0 \leq d \leq 1} \left\{\mathbb{E}[d] : \mathbb{E}[dq_b] \geq \mathbb{E}[q_b]\right\} = \max_{v \geq 0} \left\{\mathbb{E}[\min\{vq_b, 1\}]\right\}. \tag{4}$$

The minimizing tail cutoffs at $p^\dagger$ *are positive and reciprocals of the maximizing multipliers.*

3.2. The sample tipping count and the population fraction

Estimate the reference adjustments $\hat{\pi}$ and $\hat{\eta}$ once, using full-sample least squares. With $\hat{v}_i = w_i - z_i'\hat{\pi}$ and $\hat{h}_i = y_i - z_i'\hat{\eta}$, form the fitted support scores

$$\hat{q}_i := \hat{v}_i(\hat{h}_i - \hat{b}\hat{v}_i).$$

Their sum is $\sum_i \hat{q}_i = N\hat{c}\hat{L}$, where $\hat{c} = \hat{\vartheta} - \hat{b}$ and $\hat{L} := N^{-1}\sum_i \hat{v}_i^2$. Use $\hat{b} = b$ for a fixed floor and $\hat{b} = \hat{\vartheta}/2$ for the half-effect conclusion. When fitted net support is positive and at least one score is negative, a single sort per conclusion gives both the exact sample tipping count and a fractionally interpolated estimate:

Step 1. Sort $\hat{q}_{[1]} \geq \cdots \geq \hat{q}_{[N]}$, breaking ties by observation index.

Step 2. Find the first count exhausting net support:

$$k^\dagger := \min\Big\{k : \sum_{j\leq k} \hat{q}_{[j]} \geq \sum_i \hat{q}_i\Big\}.$$

These $k^\dagger$ observations form the smallest deletion that brings the slope to at most $\hat{b}$.

Step 3. Interpolate within the crossing count to obtain

$$\hat{p}^\dagger = \frac{k^\dagger - 1}{N} + \frac{\sum_i \hat{q}_i - \sum_{j<k^\dagger} \hat{q}_{[j]}}{N\hat{q}_{[k^\dagger]}}.$$

The top k scores remove the most support available from k observations. At the crossing, the negative scores remain in the sample and ensure positive identifying variation. The interpolated share $\hat{p}^\dagger$ is the unique interior zero of the empirical counterpart to (3). For $0 < p < 1$, put $k = \lfloor Np \rfloor$ and $t = Np - k$. The balance sums the top k scores and a fraction of the next, then subtracts net support, all divided by N:

$$\hat{A}_b(p) = \frac{1}{N}\left(\sum_{j=1}^{k} \hat{q}_{[j]} + t\hat{q}_{[k+1]} - \sum_{i=1}^{N} \hat{q}_i\right).$$

Rounding the interpolated estimate up recovers the exact count, $k^\dagger = \lceil N\hat{p}^\dagger \rceil$.

If $\hat{c} \le 0$, the reference estimate is already at or below the floor and we report no tipping interval for that conclusion. If net support is positive and there are no negative scores, the slope reaches the floor only by removing every strictly positive score. This yields a count if some retained zero-score observation has $\hat{v}_i \neq 0$, otherwise removing all positive support also removes all identifying variation, and no admissible deletion overturns the conclusion. The regular inference below requires some negative score in the population.

3.3. Inference for the tipping fraction

We locate the population tipping fraction by ruling out loss shares that are too small or too large. The sign of the support balance tells us whether the worst exclusion removes too little support to overturn the conclusion, just enough, or more than enough. We test the sign of that balance at candidate shares in a reporting range $\mathscr{P} = [\underline{p}, \overline{p}] \subset (0, 1)$ with $\underline{p} < \overline{p}$.

Let $O = (y, w, z)$ and let ξ_* be the cutoff separating the most supportive share $p^\dagger$ from the rest, as determined by the upper-tail problem in (2). At a regular crossing $p = p^\dagger$, the influence function distinguishes sampling variation at fixed scores from changes in the scores induced by the fitted adjustments and floor.

$$\begin{aligned}\varphi_A(O) = \underbrace{\ell_{\xi_*}(q_b) - \mathbb{E}[\ell_{\xi_*}(q_b)]}_{\text{support-tail fluctuation}} + \underbrace{a'_\eta \varphi_\eta(O)}_{\text{outcome adjustment}} \\ + \underbrace{a'_\pi \varphi_\pi(O)}_{\text{regressor adjustment}} + \underbrace{B_v \varphi_b(O)}_{\text{conclusion floor}} .\end{aligned} \tag{5}$$

The first term uses $\ell_\xi(q) = -\min\{q, \xi\}$ and captures the fluctuation in the balance for known scores. The first-order effect of estimating the cutoff cancels by optimality. In the Gaussian example with a fixed floor (Example 1), this is the only term. In an adjusted regression, the selected tail need not preserve full-population orthogonality, so the next two terms account for fitting the adjustments. The final term accounts for fitting a relative floor and vanishes

when the floor is fixed. The loadings are

$$a_\eta = -\mathbb{E}[\mathbf{1}\{q_b > \xi_*\} z v], \quad a_\pi = \mathbb{E}[\mathbf{1}\{q_b > \xi_*\} z (2bv - h)], \quad B_v = \mathbb{E}[\mathbf{1}\{q_b \le \xi_*\} v^2].$$

Sampling. The following conditions govern sampling, estimation of the adjustments and floor, and the cutoff that exhausts net support.

Assumption 1 (Sampling and regularity for tipping inference)**.**

(a) Sampling and moments. *Observations are i.i.d. draws from P and the reference projections are nonsingular. The moment condition* $\mathbb{E}[(|w| + \|z\|)^2(|y| + |w| + \|z\|)^2] < \infty$ *holds.*

(b) Estimated adjustments and floor. $(\hat{\eta}, \hat{\pi}, \hat{b})$ *is asymptotically linear with centered square-integrable joint influence, and estimates of these nuisance influence functions are consistent in empirical mean square.*

(c) Regular crossing. *Some observations oppose the conclusion, i.e.,* $P(q_b < 0) > 0$. *At* $p = p^\dagger$, *the upper-tail problem in* (2) *has a unique minimizing cutoff* ξ_*, *with* $P(q_b = \xi_*) = 0$, *and the complete influence in* (5) *has positive variance.*

For least-squares adjustments with a fixed or prespecified proportional floor, (b) follows from (a). Cutoff regularity at the crossing gives a uniform expansion in shrinking neighborhoods of $p^\dagger$ without requiring a positive score density or a cutoff rate. This lets us evaluate the balance at moving grid nodes and at the estimated crossing. Neither a linear conditional mean nor homoskedasticity is assumed. Finite-support score laws fall outside these regular-crossing conditions; Proposition S1 in the Supplement gives a local alternative for a singleton cutoff whose mass is split strictly inside it.

Theorem 1 (Asymptotic normality of the tipping estimator)**.** *Suppose the reference conclusion holds* ($c > 0$), *Assumption 1 holds, and* $p^\dagger \in (\underline{p}, \overline{p})$. *Then* $\hat{p}^\dagger$ *and* $k^\dagger/N$ *are interior to* $\mathscr{P}$ *with probability tending to one, and*

$$\sqrt{N}(\hat{p}^\dagger - p^\dagger) = -\frac{1}{\xi_* \sqrt{N}} \sum_{i=1}^{N} \varphi_A(O_i) + o_p(1), \tag{6}$$

where φ_A is the complete influence function in (5)*. Consequently,*

$$\sqrt{N}(\hat{p}^\dagger - p^\dagger) \xrightarrow{d} \mathrm{N}\Big(0, \mathrm{Var}(\varphi_A(O))/\xi_*^2\Big).$$

The integer share satisfies $0 \le k^\dagger/N - \hat{p}^\dagger < 1/N$ and has the same limit.

The proof is given in Section A.3.

In the intercept-only, fixed-floor case, the moment requirement reduces to $\mathbb{E}[y^2] < \infty$ and the adjustment and floor corrections drop out. The resulting variance $\mathrm{Var}(\min\{q_b, \xi_*\})/\xi_*^2$ is the known-score case of Mafusalov, Shapiro, and Uryasev (2018), and is positive because the score has mass on both sides of zero and $\xi_* > 0$.

The derivative $A_b'(p^\dagger) = \xi_* > 0$ is the support score at the margin of the optimal deletion. If this marginal score is small, removing another unit of population mass changes the balance only slightly and the error in the estimated balance translates into a larger error in the tipping fraction.

At each candidate share p, use the sorted fitted scores to find the cutoff and assign any fractional weight needed at a tie. Evaluate all four influence components for each observation and take the sample variance of their sum. Their covariances can increase or decrease the variance relative to treating fitted scores as known. With $\hat{\varphi}_{A,i}(p)$ denoting the sum for observation i, the variance estimate is

$$\hat{\sigma}_A^2(p) = \frac{1}{N}\sum_{i=1}^{N}\Big\{\hat{\varphi}_{A,i}(p) - \bar{\varphi}_A(p)\Big\}^2, \qquad \bar{\varphi}_A(p) = \frac{1}{N}\sum_{i=1}^{N}\hat{\varphi}_{A,i}(p).$$

Section A gives the plug-in loadings, including the fractional weight at a tied cutoff, and the least-squares influence functions for fixed and proportional floors.

3.4. Confidence intervals by support-test inversion

Consider a plausible loss budget for the reference population that lies strictly inside the reporting range $\mathscr{P}$. When a confidence interval covers the population

tipping fraction any loss budget below the lower endpoint is too small to overturn the conclusion under any admissible exclusion. Above the upper endpoint, some exclusion of that size can overturn the conclusion. Budgets falling inside the interval remain statistically unresolved. We construct this interval by dividing the reporting range into cells and testing the sign of thesupport balance at their endpoints.

Reporting range. The target parameter is the tipping fraction clipped to the reporting range, $p^{\dagger}_{\mathscr{P}} := \min\{\overline{p}, \max\{\underline{p}, p^{\dagger}\}\}$, with matching clipping applied to $\hat{p}^{\dagger}$ and $k^{\dagger}/N$. Whenever a reported confidence interval endpoint touches a boundary of $\mathscr{P}$, we flag it to indicate that the data do not rule out a crossing beyond that boundary. The lower bound $\underline{p} > 0$ excludes vanishing deletion shares from fixed-share asymptotics, while the upper bound $\overline{p} < 1$ can be chosen to reflect the largest substantively plausible attrition or contamination budget.

Let $p_{0,N} < \cdots < p_{J_N,N}$ be a deterministic grid covering $\mathscr{P}$ whose outer nodes converge to the boundaries of $\mathscr{P}$. At each grid node p, form the pointwise confidence bounds for $A_b(p)$:

$$L_A(p) = \hat{A}_b(p) - z_{1-\alpha/2}\hat{\sigma}_A(p)/\sqrt{N},$$
$$U_A(p) = \hat{A}_b(p) + z_{1-\alpha/2}\hat{\sigma}_A(p)/\sqrt{N}.$$

A cell $[p_{j,N}, p_{j+1,N}]$ is ruled out if its left endpoint lies significantly *above* zero ($L_A(p_{j,N}) > 0$) or its right endpoint lies significantly *below* zero ($U_A(p_{j+1,N}) < 0$). To ensure valid coverage when $p^{\dagger}$ lies outside $\mathscr{P}$, we omit the left test on the first cell and the right test on the last cell, so a crossing beyond $\mathscr{P}$ is retained at the nearest boundary. The reported confidence interval $[\hat{p}_L, \hat{p}_U]$ is the hull of all surviving cells clipped to $\mathscr{P}$ (defaulting to $\mathscr{P}$ itself if no cell survives). Because retaining even a single cell containing $p^{\dagger}$ suffices for coverage, and the endpoints of that cell collapse onto $p^{\dagger}$ as the grid mesh shrinks, each one-sided boundary error is asymptotically bounded by $\alpha/2$. Sections A.4 and E.1 give the boundary tests and the padded count grid used in the applications.

Theorem 2 (Tipping interval and ranges)**.** *Suppose the reference conclusion holds ($c > 0$), $P(q_b < 0) > 0$, and Assumption 1(a)–(b) hold. If $p^\dagger \in \mathscr{P}$, additionally impose the cutoff and variance conditions in (c). As the grid mesh tends to zero and the outer nodes converge to the boundaries of $\mathscr{P}$,*

$$\liminf_{N\to\infty} \mathbb{P}\left(\hat{p}_L \le p^\dagger_{\mathscr{P}} \le \hat{p}_U\right) \ge 1-\alpha.$$

On this coverage event:

(i) Preservation range. *At every share $p \in \left(\underline{p}, \hat{p}_L\right)$, every fractional exclusion preserves the conclusion;*

(ii) Overturning range. *At every share $p \in (\hat{p}_U, \overline{p})$, some fractional exclusion leaves the coefficient at or below the floor.*

The proof is given in Section A.4.

This coverage guarantee holds pointwise for a fixed population. Over an unrestricted finite-variance class, distinct populations can yield nearly indistinguishable samples and still possess different tipping fractions, precluding uniform finite-sample coverage with uniformly shrinking expected width (see Section F).

When the crossing $p^\dagger$ lies strictly outside the reporting range $\mathscr{P}$, consistency of the boundary balance estimate together with vanishing test half-widths suffices to cover the clipped target $p^\dagger_{\mathscr{P}}$. In practice, our implementation suppresses reporting an interval if the sample fails the *orientation check* ($\hat{c} > 0$, meaning the full-sample estimate does not exceed the floor being audited) or fails diagnostics on model fit, standard errors, or retained identifying variation. Counting samples as noncoverage preserves the asymptotic guarantee whenever the probability of passing all diagnostic checks approaches one.

Remark 1 (Inversion or Wald)**.** *For an interior crossing, Theorem 1 also justifies the Wald interval $\hat{p}^\dagger \pm z_{1-\alpha/2}\hat{\sigma}_A(\hat{p}^\dagger)/(\hat{\xi}\sqrt{N})$, where $\hat{\xi}$ estimates the marginal support score at $\hat{p}^\dagger$. We prefer test inversion because it handles crossings near or outside the reporting range and avoids dividing by a potentially small estimated cutoff $\hat{\xi}$.*

3.5. Extension: adverse movement at a loss budget

The tipping fraction fixes a conclusion and finds the smallest loss share that overturns it. A complementary *sensitivity curve* fixes the loss budget p and asks how far the coefficient can fall. Using the reference residual u and identifying variation L, define the normalized influence and leverage variables

$$Y := \frac{vu}{L}, \qquad Z := \frac{v^2}{L}. \tag{7}$$

The coefficient drop from an excluding d equals the deleted influence divided by the fraction of identifying variation retained:

$$\vartheta - \vartheta(1-d) = \frac{\mathbb{E}[Yd]}{\mathbb{E}[Z(1-d)]}.$$

The sensitivity curve $\delta(p)$ traces the supremum of this drop over $d \in \mathscr{D}_p$ across admissible loss budgets. On that domain, exhausting net support ($A_b p \geq 0$) is equivalent to the sensitivity curve reaching the conclusion margin ($\delta(p) \geq c$). While the tipping crossing always lies within this domain, conducting inference for one tipping fraction is considerably simpler and less demanding than estimating the full curve. It avoids constructing simultaneous confidence bands or imposing positive retained variation across the reporting range.

By contrast, the first-order AMIP curve ranks observations by the the numerator Y alone, holding the denominator $\mathbb{E}[Z(1-d)]$ fixed at one. Because $0 < \mathbb{E}[Z(1-d)] \leq 1$ on the shared domain, AMIP always produces a weakly smaller worst-case drop and can never cross a positive margin at a smaller loss budget than the exact fixed-adjustment curve. Neither curve is generally ordered relative to unrestricted refitting. The Supplement (Section S2) develops pointwise intervals at prespecified budgets and simultaneous bands for $\delta(p)$ under stronger regularity conditions than those required for the tipping interval.

3.6. Finite-sample performance

We evaluate the finite-sample coverage of our nominal 95% tipping intervals via Monte Carlo simulation across 38 primary combinations of data-generating process, sample size, and conclusion (see Table 1). Two sets of designs resample covariates and treatment assignments from our Mindspark and STAR empirical applications while varying the conditional outcome distribution. Synthetic four-cell designs compare Gaussian and heavy-tailed Student-t_3 errors across sample sizes. We also include rounded and discrete-residual outcome designs as stress tests outside the atomless-crossing assumptions.

TABLE 1: Coverage of nominal 95% tipping intervals in simulations

Simulated population	N	Coverage (%)	Lower bound (%)
Empirically calibrated designs			
Mindspark math, five laws	535	95.16–95.93	94.50
Mindspark Hindi, five laws	537	95.11–96.34	94.44
STAR, six laws	729	95.83–97.25	95.22
Stylised designs			
Normal, four-cell	200	95.59	94.96
	800	95.38	94.73
	3,200	95.13	94.47
Student-t_3, four-cell	200	94.34	93.63
	800	95.39	94.74
	3,200	95.01	94.34

Notes. Each law and sample size uses 12,000 Monte Carlo replications. Ranges span the laws and conclusions in each row. Coverage targets the crossing clipped to the reporting range, counting refusals as noncoverage. Lower bounds are the smallest lower confidence limits within each row, with 95% simultaneous confidence across all 38 coverage rates. Section E.1 describes the simulation designs.

Across all 38 primary combinations, empirical coverage ranges from 94.34% to 97.25%, reaching its lowest value under Student-t_3 errors at $N = 200$. Replications in which diagnostic checks suppress interval reporting are conservatively counted as noncoverage. The last column of Table 1 reports the smallest lower confidence limit for coverage within each row, adjusted across all 38 designs. Section E.1 gives the designs and replication counts.

In the primary synthetic experiments, nearly all unreported intervals occur in the small-sample ($N = 200$) Student-t_3 design. In separate stress tests with a small positive margin relative to sampling noise (*weak orientation*), coverage falls to 72.31–80.73% in three cases. This is driven almost entirely by samples failing the full-sample orientation check $\hat{c} > 0$ (see Section S5.2).

4. Applications

We consider Mindspark to illustrate population inference for a tipping fraction. Two sample audits — the STAR experiment and the differential effects of ruggedness on income — show how to interpret the composition of a tipping set and the identifying variation it removes.

4.1. Mindspark and uncertainty about the tipping fraction

The Mindspark experiment evaluated an after-school program combining personalized software with instructor support (Muralidharan, Singh, and Ganimian, 2019). Regressing endline math achievment on an intercept, a lottery offer indicator, and baseline math achievment yields an estimated math gain of 0.36935 baseline standard deviations for 535 students with usable scores. Of the 619 students assigned by lottery, 84 (13.6%) fall outside this regression due to a lack of usable scores. We apply that same percentage to the population represented by the 535 students with usable scores as a benchmark for *further* selective loss.

The sample gain falls by at least half after deleting 31 students (5.79%), using a floor of 0.18467 standard deviations. The nominal 95% interval for the population tipping fraction is $[3.36, 8.60]$%, entirely below the benchmark. Erasing the gain takes 77 students (14.39%), with an interpolated estimate of 14.35%. The 95% sign interval starts at 8.41% and reaches the 20% upper boundary of the prespecified reporting range $p \in [0.005, 0.20]$.[2] Thus a loss of

[2]Extending the upper boundary to 40% yields a nominal 95% interval of $[8.41, 20.93]$%.

13.6% is too small to erase the gain in the sample, but the sign interval leaves that possibility open in the population.

For population inference, we consider students who would have usable scores under the study's lottery and assessment procedures. The weight of each stratum and offer arm reflects its lottery allocation and its rate of usable assessments. We hold the released scoring map fixed and assume the crossing removes a strictly interior fraction of one type with a distinct cutoff score. With independent student records and fixed lottery counts within strata, the usual i.i.d. variance estimator is asymptotically conservative for this target (Online Supplement, Proposition S1 and corollary S1). Section D.1 gives the specification and assumptions.

TABLE 2: Mindspark sample audits and population tipping intervals

				Sample deletions		Population
Outcome	N	$\hat{\vartheta}$	SE	Exact	Interpolated	95% interval
Math	535	0.369	0.064	31 (5.79)	30.96 (5.79)	[3.36, 8.60]
Hindi	537	0.227	0.062	17 (3.17)	16.55 (3.08)	[1.12, 5.40]

Notes. Sample deletions are reported as counts with percentage shares in parentheses, relative to students with usable baseline and endline scores in each subject. The exact count is the minimum number of deletions needed to halve the adjusted gain under fixed reference adjustments. Population intervals are expressed as percentages of the observed-case target population. SE denotes the heteroskedasticity-robust (HC1) standard error of the reference coefficient. Nominal 95% intervals use paired direct-support grid-cell inversion on $[0.5, 20]$%. Confidence is asymptotic and per outcome and conclusion under the maintained models (Section S1).

Table 2 reports the half-effect results for math and Hindi. Figure 1 shows both math conclusions and their confidence intervals. The 31 students in the math half-effect set comprise 15 offered students and 16 controls, and span all four baseline-achievement quartiles. Within this set, the mean outcome residual is +1.360 among offered students and −1.647 among controls (Online Supplement, Section S6). Refitting the full regression after deleting the set gives a similar estimate.

Muralidharan, Singh, and Ganimian (2019) report Lee bounds for the math effect of a lottery offer, with a lower 95% confidence limit of 0.157 baseline

standard deviations (see their Table A.8). Under a monotonicity assumption on response, those bounds address the causal effect for students who would have usable outcomes under either assignment. Our interval asks how much further selective loss among students with usable scores could halve or erase the observed comparison.

4.2. STAR and concentration of support

The college STAR experiment offered academic support, financial incentives, both, or neither to incoming students at a Canadian university campus (Angrist, Lang, and Oreopoulos, 2009). We reproduce the women's fall-grade regression in Table 5 with all three offer indicators and the original categorical controls. The combined-offer coefficient is 4.205 grade points, with a standard error of 1.325. Of the 729 women in the regression, 62 received the combined offer. Section D.2 describes the analysis sample and controls.

The coefficient falls by at least half after deleting eight women in order of their support scores, all of whom received the combined offer. Their mean fall grade is 81.13, their mean regression residual is +14.98, and their mean high-school GPA is 78.11 compared with 79.12 overall. The full regression on the remaining sample gives 1.867 grade points, also falling below the original half-effect floor.

These eight women account for 1.10% of the regression sample and 12.9% of the combined-offer arm. The arm concentration holds across all minimum-size deletions that halve the coefficient under fixed adjustments. At least six of the eight must come from the combined-offer arm, and splits with six, seven, or eight from that arm can succeed. Restricting exclusions to the other arms raises the minimum to 31 of 667 women.

4.3. Ruggedness and the role of identifying variation

In Nunn and Puga (2012), we audit the positive Africa differential in the ruggedness–income relationship, conditional on the stated controls in Section D.3. This revisits the setting where Kuschnig, Zens, and Crespo Cuaresma

(2021) identified small influential subsets of small nations (in land area and population), and Konrad and Kuschnig (2026b) found excessive influence in an extreme-value test for influential sets.

With fixed reference adjustments, excluding ten of 170 countries suffices to reverse the sign, while the first-order AMIP approximation reaches the zero floor at 31 exclusions. The ten-country deletion leaves only 46.6% of the original identifying variation. The exact calculation divides the deleted influence by that retained fraction, while AMIP holds it fixed at unity. Refitting the full regression after deleting the same ten countries takes the interaction coefficient from +0.321 to −0.068. For this set the exact coefficient drop under fixed adjustments is about $1/0.466 = 2.15$ times its first-order drop.

5. Discussion

Two studies can report the same coefficient and standard error and still differ in how much selective loss would overturn the conclusion. Consider a model with Gaussian errors and one with scaled logistic errors that preserves the coefficient and asymptotic variance. The sign tips earlier under logistic errors (see Section S4.2 in the Supplement). The difference comes from how much support the most favorable observations contribute — a feature of the outcome distribution that the standard error leaves undetermined.

5.1. Information to guide exclusion

Consider a randomized offer D with common assignment probability e, independent of baseline characteristics X and both potential outcomes. Let the controls consist of an intercept and functions of X. If loss can depend only on X, the most support that can be removed at share p is $\mathrm{T}_p(\mathbb{E}[q_b \mid X])$ by (11). Since $v = D - e$,

$$\mathbb{E}[q_b \mid X] = e(1-e)\{\tau(X) - b\}, \qquad \tau(X) = \mathbb{E}[Y(1) - Y(0) \mid X].$$

The most adverse baseline-only rule removes students with the largest conditional treatment effects, and the retained slope equals the average treatment effect among those who remain (see Remark 3). This is the connection to worst-case effects over covariate-defined subpopulations (Jeong and Namkoong, 2020).

Now suppose that assignment is balanced, $Y(0) = \varepsilon$ and $Y(1) = \tau + \varepsilon$, with $\varepsilon \sim \mathrm{N}(0, \sigma^2)$ independent of X and D. Every student gains the same amount, so every baseline-defined group retains the effect τ. Yet high outcomes among offered students and low outcomes among controls both support a positive adjusted comparison. With an intercept as the only control and $b = 0$, their support score is $q_0 = \tau/4 + (D - 1/2)\varepsilon$. At $\tau = 0.2\sigma$, its mean is one tenth of its standard deviation. The calculation in Example 1 then gives a tipping fraction of about 4.5%. Selective loss can erase the comparison even when every student benefits equally.

5.2. Practical guidance

The audit is most informative when anchored to a concrete empirical question rather than reported as standalone diagnostic. In applied work, a researcher would start with the conclusion of interest, the population from which records could be lost, and a plausible loss budget. She would report the coefficient and standard error, the exact count $k^\dagger$ out of N, the interpolated population estimate $\hat{p}^\dagger$, and its interval (flagging any reporting boundary). For a budget inside the reporting range, a value below the lower endpoint supports preservation under every admissible exclusion. Above the upper endpoint, some exclusion of that size can overturn the conclusion. A budget inside the interval remains inconclusive. The exclusion that can overturn the conclusion is a worst-case and may still be unlikely in practice. Joint coverage across conclusions requires a multiplicity adjustment or joint inference.

The selected observations help judge the loss mechanism. In STAR the eight women in the highest-scoring set account for roughly one in eight recipients of the combined offer. If records could be lost only from the other arms, the

minimum rises to 31. For Mindspark, halving the math gain takes 46 of 264 offered students under offer-only loss, or 41 of 271 controls under control-only loss (see Section S6). Where the sampling design supports only a sample audit, the exact count and the composition of an attaining set can stand on their own.

5.3. Checking the selected set by refitting

The common adjustment leaves each retained observation at its full-sample coordinates in the added-variable plot. Refitting the full regression lets the remaining observations redetermine those coordinates. The two slopes agree if the adjusted regressor remains orthogonal to the controls after exclusion. Otherwise the gap can go in either direction, increasing in the explanatory power of the controls for the remaining regressor variation.

As long as the retained design maintains full rank, refitting after deleting the selected set is a simple crosscheck. In the ruggedness application, for example, deleting the ten selected countries yields an unrestricted refit of -0.068. This establishes that ten deletions suffice under refitting, although a smaller set might also suffice.

5.4. Scope of population inference

The tipping interval has asymptotic coverage for each fixed population satisfying the stated conditions. Over an unrestricted finite-variance class, populations that are difficult to distinguish in a sample can have different tipping fractions (Section F). A small positive reference margin poses a practical difficulty even in the Gaussian designs. The sample coefficient can fall below the floor before any records are removed. Our implementation returns no interval in this case, and the calibration counts the sample as noncoverage (Section S5.2).

For recorded outcomes, what matters is how the observations at the cutoff respond to fitting the scores. The Mindspark result allows repeated copies of one record type at the cutoff when tipping removes a strictly interior fraction

of that type. Distinct types tied at the cutoff can change rank when the adjustment is fitted. The finite-type result also relies on the sampling model for its standard error. Extending inference to shared classroom shocks, for example, would require accounting for dependence among students even if the audit continued to exclude individual records. Developing population intervals for restrictions on who can be lost, such as loss confined to one arm, is a further question suggested by the sample audits.

6. Conclusion

A tipping fraction attaches a loss budget to a particular claim about a coefficient. In Mindspark, 31 of 535 assessed students suffice to halve the estimated math gain under the original adjustment. Under the observed-case population model, the 13.6% benchmark for further selective loss exceeds the entire interval for halving the gain. The interval for erasing the gain includes that same budget. For a decision that requires more than half the reference gain, loss of this size could overturn the conclusion. For a decision that requires only a positive gain, the evidence leaves that possibility open.

References

Andrews, I., M. Gentzkow, and J. M. Shapiro (2017). “Measuring the Sensitivity of Parameter Estimates to Estimation Moments”. In: *The Quarterly Journal of Economics* 132.4, pp. 1553–1592. DOI: 10.1093/qje/qjx023.

Angrist, J. D., D. Lang, and P. Oreopoulos (2009). “Incentives and Services for College Achievement: Evidence from a Randomized Trial”. In: *American Economic Journal: Applied Economics* 1.1, pp. 136–163. DOI: 10.1257/app.1.1.136.

Armstrong, T. B. and M. Kolesár (2021). “Sensitivity Analysis Using Approximate Moment Condition Models”. In: *Quantitative Economics* 12.1, pp. 77–108. DOI: 10.3982/qe1609.

Berrendero, J. R. and R. H. Zamar (2001). "Maximum Bias Curves for Robust Regression with Non-Elliptical Regressors". In: *The Annals of Statistics* 29.1, pp. 224–251. DOI: 10.1214/aos/996986507.

Bonhomme, S. and M. Weidner (2022). "Minimizing Sensitivity to Model Misspecification". In: *Quantitative Economics* 13.3, pp. 907–954. DOI: 10.3982/qe1930.

Chatterjee, S. and A. S. Hadi (1986). "Influential Observations, High Leverage Points, and Outliers in Linear Regression". In: *Statistical Science* 1.3, pp. 379–393. DOI: 10.1214/ss/1177013622.

Christensen, T. and B. Connault (2023). "Counterfactual Sensitivity and Robustness". In: *Econometrica* 91.1, pp. 263–298. DOI: 10.3982/ecta17232.

Cook, R. D. (1979). "Influential Observations in Linear Regression". In: *Journal of the American Statistical Association* 74.365, pp. 169–174. DOI: 10.2307/2286747.

Dewaskar, M. et al. (2025). "Robustifying Likelihoods by Optimistically Re-weighting Data". In: *Journal of the American Statistical Association* 120.551, pp. 1787–1798. DOI: 10.1080/01621459.2025.2468012.

Donoho, D. L. and P. J. Huber (1983). "The Notion of Breakdown Point". In: *A Festschrift for Erich L. Lehmann*. Ed. by P. J. Bickel, K. A. Doksum, and J. L. Hodges Jr. Belmont, CA: Wadsworth, pp. 157–184.

Giordano, R., R. Meager, and T. Broderick (2026a). "An Automatic Finite-Sample Robustness Metric: When Can Dropping a Little Data Change Conclusions? Part I: Definitions and Experiments". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 384.2321, p. 20250001. DOI: 10.1098/rsta.2025.0001.

– (2026b). "An Automatic Finite-Sample Robustness Metric: When Can Dropping a Little Data Change Conclusions? Part II: Theory and Intuition". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 384.2321, p. 20240614. DOI: 10.1098/rsta.2024.0614.

Hampel, F. R. (1974). "The Influence Curve and Its Role in Robust Estimation". In: *Journal of the American Statistical Association* 69.346, pp. 383–393. DOI: 10.1080/01621459.1974.10482962.

Horowitz, J. L. and C. F. Manski (1995). "Identification and Robustness with Contaminated and Corrupted Data". In: *Econometrica* 63.2, p. 281. DOI: 10.2307/2951627.

Hu, Y. et al. (2024). "Most Influential Subset Selection: Challenges, Promises, and Beyond". In: *Conference on Neural Information Processing Systems (NeurIPS 2024)*. Vol. 38. DOI: 10.48550/arxiv.2409.18153.

Huber, P. J. (1981). *Robust Statistics*. 1st ed. Wiley Series in Probability and Statistics. Wiley. DOI: 10.1002/0471725250.

Ichimura, H. and W. K. Newey (2022). "The Influence Function of Semiparametric Estimators". In: *Quantitative Economics* 13.1, pp. 29–61. DOI: 10.3982/qe826.

Jeong, S. and H. Namkoong (2020). "Robust causal inference under covariate shift via worst-case subpopulation treatment effects". In: *Proceedings of Thirty Third Conference on Learning Theory*. 33rd Annual Conference on Learning Theory. Vol. 125. Proceedings of Machine Learning Research. PMLR, pp. 2079–2084.

Kitamura, Y., T. Otsu, and K. Evdokimov (2013). "Robustness, Infinitesimal Neighborhoods, and Moment Restrictions". In: *Econometrica* 81.3, pp. 1185–1201. DOI: 10.3982/ecta8617.

Kline, P. and A. Santos (2013). "Sensitivity to Missing Data Assumptions: Theory and an Evaluation of the U.S. Wage Structure: Sensitivity to Missing Data Assumptions". In: *Quantitative Economics* 4.2, pp. 231–267. DOI: 10.3982/qe176.

Koh, P. W. and P. Liang (2017). "Understanding Black-box Predictions via Influence Functions". In: *Proceedings of the 34th International Conference on Machine Learning*. Vol. 70. Proceedings of Machine Learning Research. PMLR, pp. 1885–1894. DOI: 10.48550/arXiv.1703.04730.

Konrad, L. D. and N. Kuschnig (2026a). "Finding Most Influential Sets". In: *Proceedings of the 43rd International Conference on Machine Learning (ICML)*. Vol. 306. Proceedings of Machine Learning Research. PMLR, pp. 60188–60203. DOI: 10.48550/arxiv.2606.05919.

– (2026b). "Testing Most Influential Sets". In: *Proceedings of the 14th International Conference on Learning Representations (ICLR)*. DOI: 10.48550/arxiv.2510.20372.

Kuschnig, N., G. Zens, and J. Crespo Cuaresma (2021). "Hidden in Plain Sight: Influential Sets in Linear Regression". In: *CESifo Working Paper* 8981. DOI: 10.13140/rg.2.2.30682.47042.

Mafusalov, A., A. Shapiro, and S. Uryasev (2018). "Estimation and Asymptotics for Buffered Probability of Exceedance". In: *European Journal of Operational Research* 270.3, pp. 826–836. DOI: 10.1016/j.ejor.2018.01.021.

Masten, M. A. and A. Poirier (2020). "Inference on Breakdown Frontiers". In: *Quantitative Economics* 11.1, pp. 41–111. DOI: 10.3982/qe1288.

Moitra, A. and D. Rohatgi (2023). "Provably Auditing Ordinary Least Squares in Low Dimensions". In: *International Conference on Learning Representations (ICLR 2023)*. DOI: 10.48550/arxiv.2205.14284.

Muralidharan, K., A. Singh, and A. J. Ganimian (2019). "Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India". In: *American Economic Review* 109.4, pp. 1426–1460. DOI: 10.1257/aer.20171112.

Nunn, N. and D. Puga (2012). "Ruggedness: The Blessing of Bad Geography in Africa". In: *Review of Economics and Statistics* 94.1, pp. 20–36. DOI: 10.1162/rest_a_00161.

Ober-Reynolds, D. (2026). "Robustness to Missing Data: Breakdown Point Analysis". In: *Journal of Econometrics* 253, 106151, p. 106151. DOI: 10.1016/j.jeconom.2025.106151.

Palomba, F. (2026). *Sensitivity Analysis in Population Shares*. Version 1. DOI: 10.48550/arxiv.2611.05584. URL: https://arxiv.org/abs/2611.05584v1 (visited on 11/06/2026).

Rockafellar, R. T. and S. Uryasev (2000). "Optimization of Conditional Value-at-Risk". In: *The Journal of Risk* 2.3, pp. 21–41. DOI: 10.21314/JOR.2000.038.

Rubinstein, I. and S. B. Hopkins (2024). "Robustness Auditing for Linear Regression: To Singularity and Beyond". In: *ArXiv e-prints*. DOI: 10.48550/arxiv.2410.07916.

Spini, P. E. (2026). "A Simple Measure of Robustness for External Validity under Covariate Shifts". In: *Econometrica*. Forthcoming.

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.

A. Support-Test Inference for the Tipping Fraction

We prove the tipping characterization, derive a local expansion of the fitted support balance without a cutoff density assumption, and use it for the tipping estimator and grid-cell interval.

A.1. Support characterization and minimum exclusion

Use the coefficient notation of Section 2, with $q_b = v(h - bv)$ and $c = \vartheta - b$.

Lemma 1 (Concavity of the tail integral)**.** *Let V be integrable. Then $p \mapsto \mathrm{T}_p(V)$ is concave and absolutely continuous on $[0,1]$, with $\mathrm{T}_0(V) = 0$, $\mathrm{T}_1(V) = \mathbb{E}[V]$ and derivative $F_V^{-1}(1-p)$ almost everywhere. Consequently:*

(i) *if $\mathbb{E}[V] = 0$ and V is nondegenerate, then $\mathrm{T}_p(V) > 0$ for every $p \in (0,1)$;*

(ii) *if $\mathbb{E}[|q_b|] < \infty$ and $\mathbb{E}[q_b] > 0$, then A_b is concave with $A_b(0) < 0 = A_b(1)$. Writing $p_A^\dagger := \inf\{p \in [0,1] : A_b(p) \geq 0\}$, the balance is negative on $[0, p_A^\dagger)$ and nonnegative on $[p_A^\dagger, 1]$, and strictly positive on $(p_A^\dagger, 1)$ when $P(q_b < 0) > 0$.*

Proof of Lemma 1. Integrability of the quantile gives absolute continuity and the endpoint values. Its almost-everywhere derivative $F_V^{-1}(1-p)$ is nonincreasing, giving concavity. For (i), a concave function with zero endpoints is nonnegative, and an interior zero forces it to vanish everywhere by the chord inequalities. Its derivative would then give $V = 0$ almost surely, contradicting nondegeneracy. For (ii), the closed superlevel set $\{p \in [0,1] : A_b(p) \geq 0\}$ is an interval containing 1 and excluding 0, so it equals $[p_A^\dagger, 1]$ with $A_b(p_A^\dagger) = 0$. If $P(q_b < 0) > 0$, then $p_+ := P(q_b > 0) < 1$ and $\mathrm{T}_{p_+}(q_b) = \mathbb{E}[(q_b)_+]$, so $A_b(p_+) = \mathbb{E}[(-q_b)_+] > 0$ and hence $p_A^\dagger < p_+$. The chords from $p_A^\dagger$ to p_+ and from p_+ to 1 give strict positivity on $(p_A^\dagger, 1)$. □

Proof of Proposition 1. Primal. Write $m = \mathbb{E}[q_b]$ and let $p_A^\dagger$ be the first zero of A_b. By Lemma 1, $0 < p_A^\dagger < P(q_b > 0)$ and the balance is strictly signed on either side for $p < 1$. Below that zero, (2) gives $\mathbb{E}[(1-d)q_b] > 0$ for

every $d \in \mathscr{D}_p$, which implies positive retained variation since $q_b = 0$ wherever $v = 0$. At and above the zero for $p < 1$, upper-tail exclusion leaves support $-A_b(p) \leq 0$ and retains some negative scores (hence positive variation); at the zero it deletes only positive scores. Thus (1) identifies $p_A^\dagger$ with $p^\dagger$ and gives the two sign ranges. By (2), $p_A^\dagger$ is also the smallest deleted mass satisfying $\mathbb{E}[dq_b] \geq m$, as stated in (4).

Dual. At $p^\dagger$, (2) gives $m = \min_\xi \{\mathbb{E}[(q_b - \xi)_+] + p^\dagger \xi\}$. Every minimizing ξ is positive: if $\xi \leq 0$, the mass strictly above it exceeds $p^\dagger$ since $p^\dagger < P(q_b > 0)$. Thus for any $\xi > 0$,

$$\mathbb{E}[\min\{q_b/\xi, 1\}] = \frac{m - \mathbb{E}[(q_b - \xi)_+]}{\xi} \leq p^\dagger,$$

with equality exactly at the minimizing cutoffs. The substitution $v = 1/\xi$ gives the dual and the reciprocal correspondence, since $v = 0$ gives value zero. □

Verification of the sample tipping count. The empirical version of (2) gives the threshold form of the support balance:

$$\hat{A}_b(p) = \min_{\xi \in \mathbb{R}} \left\{ \frac{1}{N} \sum_{i=1}^{N} (\hat{q}_i - \xi)_+ + p\xi \right\} - \frac{1}{N} \sum_{i=1}^{N} \hat{q}_i. \tag{8}$$

For $0 < p < 1$ and $k = \lfloor Np \rfloor$, the counts strictly above and at or above $\hat{q}_{[k+1]}$ bracket Np, even with ties, so evaluating at this cutoff gives the sorted-sum form in Section 3.2. At a count node a literal top-k subset attains the tail maximum. With positive net support and at least one negative score, the first crossing score is positive, so the negative scores remain and ensure positive retained variation. The balance is negative at $(k^\dagger - 1)/N$ and nonnegative at $k^\dagger/N$, giving the ceiling identity by linear interpolation.

A.2. Support expansion with fitted scores

We first measure sampling variation with known support scores, then account for how fitting the adjustments and floor changes the balance.

Let $D_p = \mathbf{1}\{q_b > \xi_p\} + \gamma_p \mathbf{1}\{q_b = \xi_p\}$, where ξ_p is the upper-p support-score cutoff and $\gamma_p \in [0,1]$ satisfies $\mathbb{E}[D_p] = p$. At the crossing $p = p^\dagger$, Assumption 1(c) rules out a cutoff atom, so the fractional term vanishes almost surely and $D_{p^\dagger} = \mathbf{1}\{q_b > \xi_*\}$. Define

$$\ell_{\xi_p}(q_b) = (q_b - \xi_p)_+ - q_b, \qquad a_\eta(p) = -\mathbb{E}[D_p z v],$$
$$a_\pi(p) = \mathbb{E}[D_p z(2bv - h)], \qquad B_v(p) = \mathbb{E}[(1 - D_p)v^2],$$

which specialize at $p = p^\dagger$ to the quantities in (5).

A.2.1. A shared empirical-process bound

Let P_N be the empirical average and $\mathbb{G}_N = \sqrt{N}(P_N - P)$ the centered empirical process under law P.

Lemma 2 (Compact parameter meshes)**.** *Let $\Theta \subset \mathbb{R}^m$ be compact and $(\theta, o) \mapsto h_\theta(o)$ jointly measurable, with*

$$|h_\theta| \le M, \qquad |h_\theta - h_{\theta'}| \le M \left\|\theta - \theta'\right\|, \qquad M \in L^2(P).$$

Then $\{h_\theta : \theta \in \Theta\}$ is pointwise measurable and P-Donsker, with polynomial $L^2(P)$ bracketing numbers. For a constant C_Θ, its empirical covering numbers satisfy, simultaneously for every $\varepsilon > 0$,

$$\log \mathcal{N}(\varepsilon, \{h_\theta : \theta \in \Theta\}, \|\cdot\|_{N,2}) \le m \log\left(1 + C_\Theta \|M\|_{N,2} / \varepsilon\right), \qquad \|h\|_{N,2}^2 := P_N h^2.$$

Squares and pairwise products of this family obey uniform laws of large numbers.

Proof. Restricting to a countable dense subset of Θ gives pointwise measurability by the Lipschitz inequality. Euclidean nets of radius ε yield brackets $h_\theta \pm \varepsilon M$, giving polynomial bracketing and the Donsker conclusion (van der Vaart and Wellner, 1996, Theorem 2.7.11). Independently, nets of radius $\varepsilon / \|M\|_{N,2}$ give the displayed empirical covers; if $\|M\|_{N,2} = 0$, the empirical class is a singleton. The continuous image of the compact set Θ in $\mathbb{R}^N$ is a

measurable random compact set, so these empirical covering numbers are measurable. For pairwise products, the same finite nets have errors bounded by a constant multiple of εM^2, and the law of large numbers with $PM^2 < \infty$ yields their uniform laws. □

A.2.2. Local expansion and fitted variance

Proposition 2 (Direct support-balance test at the crossing)**.** *Suppose Assumption 1 holds, and let $p = p^\dagger \in (0,1)$ be its scalar crossing, with $\xi_p = \xi_*$. Let $(\varphi_\eta, \varphi_\pi, \varphi_b)$ be the joint nuisance influence. Then*

$$\sqrt{N}\{\hat{A}_b(p) - A_b(p)\} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \varphi_{A,i}(p) + o_p(1),$$
$$\varphi_{A,i}(p) = \ell_{\xi_p}(q_{b,i}) - \mathbb{E}[\ell_{\xi_p}(q_b)] + a_\eta(p)'\varphi_{\eta,i} + a_\pi(p)'\varphi_{\pi,i} + B_v(p)\varphi_{b,i}. \tag{9}$$

For ordinary least squares, let $Q_z := \mathbb{E}[zz']$; then $\varphi_\eta = Q_z^{-1} zh$, $\varphi_\pi = Q_z^{-1} zv$, and $\varphi_b = 0$ for a fixed floor or $\varphi_b = \kappa vu/L$ for $b = \kappa\vartheta$ with prespecified $0 \le \kappa < 1$. Empirical substitution and centering give $\hat{\sigma}_A^2(p) \xrightarrow{p} \mathrm{Var}\{\varphi_A(p)\}$, with positive limiting variance by (c). More strongly, for every deterministic $\delta_N \downarrow 0$,

$$\sup_{|r-p| \le \delta_N} \left| \sqrt{N}\{\hat{A}_b(r) - A_b(r)\} - N^{-1/2} \sum_i \varphi_{A,i}(p) \right| = o_p(1). \tag{10}$$

The plug-in variance $\hat{\sigma}_A^2(r)$ converges uniformly on this shrinking neighborhood to the same limit. Consequently, every deterministic sequence $p_N \to p$ has the centered studentized standard normal limit needed at the neighboring grid endpoints.

Proof. Write $q = q_b$ and let $\theta_{\mathrm{nu}} = (\eta, \pi, b) \in \Theta$ collect the nuisance parameters, with estimate $\hat{\theta}_{\mathrm{nu}} = (\hat{\eta}, \hat{\pi}, \hat{b})$ and error $\Delta = \hat{\theta}_{\mathrm{nu}} - \theta_{\mathrm{nu}}$. The envelope $\mathscr{H} := (|w| + \|z\|)(|y| + |w| + \|z\|) \in L^2(P)$ bounds the score and its first and second derivatives locally around θ_{nu}, up to a constant. For least squares, the normal equations give $\hat{Q}_z(\hat{\eta} - \eta) = P_N[zh]$, $\hat{Q}_z(\hat{\pi} - \pi) = P_N[zv]$, and $\hat{L}(\hat{\vartheta} - \vartheta) = P_N[\hat{v}u]$, where $\hat{Q}_z = P_N[zz']$. Since $\hat{Q}_z \xrightarrow{p} Q_z$ and $P_N[zu] = O_p(N^{-1/2})$, the

adjustment expansions follow and $P_N[\hat{v}u] = P_N[vu] + O_p(N^{-1})$; also $\hat{L} \xrightarrow{p} L$. Each plug-in influence differs from its population counterpart by at most $o_p(1)\mathscr{H}$ uniformly over observations, by consistency of these moments and the projection coefficients; $P_N\mathscr{H}^2 = O_p(1)$ then verifies (b). For population cutoffs use the measurable convention $\xi_r = F_q^{-1}(1-r)$. Sample cutoffs and fractional weights use the descending ordering in Section 3.2. The bounds hold for every minimizer on each sample path, so the localization requires no separate measurable selection.

Localization without a cutoff rate. The cutoff condition $P(q > \xi) \le p \le P(q \ge \xi)$ and atomlessness give $P(q > \xi_*) = p$. For every sufficiently small fixed $\varepsilon > 0$,

$$P(q > \xi_* - \varepsilon) > p > P(q \ge \xi_* + \varepsilon).$$

If either inequality failed, the corresponding adjacent threshold would also satisfy the cutoff condition, contradicting uniqueness. These strict count gaps persist for $|r - p| \le \delta_N$ and shrinking nuisance neighborhoods. Indeed, on $\mathscr{H} \le K$ score perturbations are bounded by $CK\|\Delta\|$, while the remaining count is bounded by the tail frequency of $\mathscr{H}$. Applying the law of large numbers to fixed indicators and then letting $K \to \infty$ traps all empirical minimizing cutoffs for the true and fitted scores, as well as all population minimizing cutoffs, within ε of ξ_* with probability tending to one. No quantitative lower bound on the count gaps is needed; neighboring cutoffs may be nonunique or atomic.

Sampling variation with known scores. Put $A_N^0(r) = \inf_\xi\{P_N\ell_\xi(q) + r\xi\}$. For any oracle empirical minimizer $\hat{\xi}_r^0$, optimality gives

$$\mathbb{G}_N\ell_{\hat{\xi}_r^0}(q) \le \sqrt{N}\{A_N^0(r) - A_b(r)\} \le \mathbb{G}_N\ell_{\xi_r}(q).$$

On a fixed compact interval around ξ_*, the hinge class is P-Donsker by Lemma 2. Localization holds for every fixed $\varepsilon > 0$, so the maximal cutoff distance is $o_p(1)$, as is the maximal $L^2(P)$ distance by $\left\|\ell_\xi - \ell_{\xi_*}\right\|_{L^2(P)} \le |\xi - \xi_*|$. Asymptotic $L^2(P)$ equicontinuity then makes both bounds $\mathbb{G}_N\ell_{\xi_*}(q) + o_p(1)$ uniformly over $|r - p| \le \delta_N$. The constants $r\xi$ cancel on centering; neither a squared-cutoff remainder nor a cutoff convergence rate enters.

Effect of fitting the scores. Let s_r^0, s_r^1 be oracle and fitted fractional maximizing weights, each with empirical mass r, and put $D_* = \mathbf{1}\{q > \xi_*\}$. For fixed $\varepsilon, K > 0$, localization and $CK\|\Delta\| \le \varepsilon$ confine disagreement on $\{\mathscr{H} \le K\}$ to $\{|q - \xi_*| \le 2\varepsilon\}$ for both sets of weights, uniformly on the shrinking budget neighborhood. The weighted disagreement is thus bounded by the empirical envelope mass in that band plus $P_N[\mathscr{H}\mathbf{1}\{\mathscr{H} > K\}]$ with probability tending to one. Passing to the limit as $N \to \infty$ via the law of large numbers, and then letting $\varepsilon \downarrow 0$ and $K \uparrow \infty$, yields $P_N[\mathscr{H}|s_r^j - D_*|] = o_p(1)$ for $j = 0, 1$ by cutoff atomlessness and integrability of $\mathscr{H}$. Taylor expansion gives $q_{\hat\theta_{\mathrm{nu}}} - q = \dot q_0'\Delta + O(\mathscr{H}\|\Delta\|^2)$. Maximizing-value optimality bounds $\hat A_b(r) - A_N^0(r)$ between $P_N[(s_r^j - 1)(q_{\hat\theta_{\mathrm{nu}}} - q)]$ for $j = 0, 1$, hence, uniformly,

$$\hat A_b(r) - A_N^0(r) = d_*'\Delta + o_p(1)\|\Delta\| + O_p(1)\|\Delta\|^2, \qquad d_* = \mathbb{E}[(D_* - 1)\dot q_0].$$

Substituting the joint nuisance expansion proves (10) with influence $\ell_{\xi_*}(q) - \mathbb{E}[\ell_{\xi_*}(q)] + d_*'\varphi_{\mathrm{nu}}$, where $\varphi_{\mathrm{nu}} = (\varphi_\eta', \varphi_\pi', \varphi_b)'$. The derivatives $\dot q_\eta = -zv$, $\dot q_\pi = z(2bv - h)$ and $\dot q_b = -v^2$, together with $\mathbb{E}[zv] = \mathbb{E}[zh] = 0$, identify $d_* = (a_\eta', a_\pi', B_v)'$ and give (9).

Fitted variance. With $\hat d_r = P_N[(s_r^1 - 1)\dot q_{\hat\theta_{\mathrm{nu}}}]$, the fitted influence and its centered empirical variance are

$$\hat\varphi_{A,i}(r) = \ell_{\hat\xi_r}(\hat q_i) + \hat d_r'\hat\varphi_{\mathrm{nu},i}, \qquad \hat\sigma_A^2(r) = P_N[(\hat\varphi_A(r) - P_N[\hat\varphi_A(r)])^2].$$

For ordinary least squares, use $\hat\varphi_{\eta,i} = \hat Q_z^{-1} z_i \hat h_i$ and $\hat\varphi_{\pi,i} = \hat Q_z^{-1} z_i \hat v_i$. Use $\hat\varphi_{b,i} = 0$ for a fixed floor and $\hat\varphi_{b,i} = \kappa \hat v_i \hat u_i / \hat L$ for a relative floor $b = \kappa\vartheta$. The weighted consistency just proved gives $\hat d_r \to d_*$ uniformly, and hinge contraction gives

$$P_N|\ell_{\hat\xi_r}(q_{\hat\theta_{\mathrm{nu}}}) - \ell_{\xi_*}(q)|^2 \le C\{\|\Delta\|^2 P_N\mathscr{H}^2 + |\hat\xi_r - \xi_*|^2\} = o_p(1).$$

The empirical-L^2 condition (b) gives $P_N\|\hat\varphi_{\mathrm{nu}}\|^2 = O_p(1)$ and empirical-L^2 convergence of the nuisance term. For OLS, $P_N[z_i\hat v_i] = P_N[z_i\hat h_i] = 0$, so $\hat d_r$ coincides with the selected-moment plug-in $(\hat a_\eta(r)', \hat a_\pi(r)', \hat B_v(r))'$ obtained by replacing D_r with s_r^1. Centering, the second-moment law and Cauchy–Schwarz give

uniform variance consistency. Together with the local expansion and positive limiting variance, this consistency justifies studentization at deterministic shares converging to the crossing. □

For least squares with a fixed or prespecified proportional floor, conditional full support of y given (w, z) suffices for positive variance at the crossing. Wherever $v \neq 0$, the known-score term has distinct slopes in y on either side of the cutoff, while the nuisance corrections are affine in y. Full support therefore gives positive conditional variance on $\{v \neq 0\}$, an event of positive probability since $L > 0$.

A.3. Asymptotic distribution of the tipping estimator

Proof of Theorem 1. Slope. The superdifferential of A_b at $p^\dagger$ consists of the minimizing cutoffs in (2), so (c) makes it the singleton $\{\xi_*\}$. Thus A_b is differentiable at $p^\dagger$ with $A_b'(p^\dagger) = \xi_* > 0$, without a density assumption.

Rate. Let $\hat{F}_{q_b}(t) = N^{-1}\sum_i \mathbf{1}\{\hat{q}_i \le t\}$ and let $\hat{F}_{q_b}^{-1}$ be its left-continuous quantile. The fractional balance $\hat{A}_b(p) = \int_0^p \hat{F}_{q_b}^{-1}(1-s)\,ds - P_N\hat{q}$ is concave, so its nonnegative set is an interval. Fix $\varepsilon > 0$ and put $p_N^\pm = p^\dagger \pm M/\sqrt{N}$. Apply (10) with $\delta_N = M/\sqrt{N}$ and subtract its value at $p^\dagger$ to cancel the common influence uniformly over $|p - p^\dagger| \le M/\sqrt{N}$. Differentiability of A_b at $p^\dagger$ then gives

$$\sqrt{N}\hat{A}_b(p_N^\pm) = \pm\xi_* M + \sqrt{N}\hat{A}_b(p^\dagger) + o_p(1),$$

where $\sqrt{N}\hat{A}_b(p^\dagger) = O_p(1)$ by Proposition 2. Choose M large enough that $\mathbb{P}\{|\sqrt{N}\hat{A}_b(p^\dagger)| < \xi_* M/2\} \ge 1 - \varepsilon/2$ for all large N. Then $\hat{A}_b(p_N^-) < 0 < \hat{A}_b(p_N^+)$ with probability at least $1-\varepsilon$ for all large N. Concavity places the first zero $\hat{p}^\dagger$ between these points, proving $\sqrt{N}(\hat{p}^\dagger - p^\dagger) = O_p(1)$ and an interior crossing with probability tending to one.

Limit. Since $A_b(p^\dagger) = 0$, Proposition 2 applies directly. Write $h_N = \sqrt{N}(\hat{p}^\dagger - p^\dagger)$. On $|h_N| \le M$, the local uniform expansion gives

$$0 = \sqrt{N}\hat{A}_b(p^\dagger) + \sqrt{N}\{A_b(\hat{p}^\dagger) - A_b(p^\dagger)\} + o_p(1) = \sqrt{N}\hat{A}_b(p^\dagger) + \xi_* h_N + o_p(1).$$

Tightness permits $M \to \infty$ after N, hence $h_N = -\sqrt{N}\hat{A}_b(p^\dagger)/\xi_* + o_p(1)$. The support-balance CLT proves (6) without ratio assumptions.

Integer crossing. Score consistency and the law of large numbers give a positive mean of the fitted support scores and at least one negative score with probability tending to one. The ceiling identity of Section 3.2 yields an admissible count with $0 \le k^\dagger/N - \hat{p}^\dagger < 1/N$, so interiority and the asymptotic distribution transfer.

Cutoff and variance consistency. Take $\delta_N = N^{-1/4}$, so $\mathbb{P}\{|\hat{p}^\dagger - p^\dagger| \le \delta_N\} \to 1$. The all-minimizer localization in Proposition 2 traps fitted minimizing cutoffs uniformly on this neighborhood, including the marginal crossing order statistic $\hat{\xi}$, near ξ_*. Its local uniform variance consistency likewise applies at $\hat{p}^\dagger$. Therefore $\hat{\xi} \xrightarrow{p} \xi_*$ and $\hat{\sigma}_A^2(\hat{p}^\dagger) \xrightarrow{p} \mathrm{Var}(\varphi_A)$; Slutsky's theorem justifies the Wald alternative of Remark 1. □

A.4. Grid-cell coverage

It suffices to retain one cell containing the crossing, using the nearest end cell when the crossing falls outside the grid. Apply the rule of Section 3.4 on the deterministic grid $p_{0,N} < \cdots < p_{J_N,N}$ with bounds L_M, U_M for a generic criterion M.

Corollary 1 (Grid-cell coverage)**.** *Suppose $M(p) < 0$ for $p < p^\dagger$ and $M(p) \ge 0$ for $p > p^\dagger$ on the grid domain, with $M(p^\dagger) = 0$ when $p^\dagger$ belongs to that domain. Choose a deterministic cell containing* $\min\{p_{J_N,N}, \max\{p_{0,N}, p^\dagger\}\}$ *(either neighboring cell at a grid node). At its nonomitted left endpoint l_N and right endpoint r_N, suppose respectively*

$$\limsup_N \mathbb{P}\{L_M(l_N) > 0\} \le \alpha/2, \qquad \limsup_N \mathbb{P}\{U_M(r_N) < 0\} \le \alpha/2.$$

Then the clipped hull $[\hat{p}_L, \hat{p}_U]$ covers $p^\dagger_{\mathscr{P}}$ with asymptotic probability at least $1-\alpha$, including crossings outside the grid or reporting range. Consistent estimators of the asymptotic standard deviations and centered standard normal limits at the tested endpoint sequences suffice for the error bounds. For a fixed finite grid,

pointwise limits at its tested endpoints suffice; a changing count grid requires sequential limits, not merely limits at each fixed share.

Proof of Corollary 1. Adjacent tests and clipping. Let C_N be the specified cell containing the grid-clipped crossing $\min\{p_{J_N,N}, \max\{p_{0,N}, p^\dagger\}\}$. Every nonomitted left test has $M(l_N) \le 0$ and every nonomitted right test has $M(r_N) \ge 0$. If $p^\dagger < p_{0,N}$, the first cell is tested only at its right endpoint; if $p^\dagger > p_{J_N,N}$, the last cell is tested only at its left endpoint. The union bound ensures that C_N is accepted with liminf probability at least $1-\alpha$, regardless of the number of grid cells. The hull of the accepted cells then contains the grid-clipped crossing, and coordinate-wise clipping to $[\underline{p}, \overline{p}]$ transfers this inclusion to coverage of $p^\dagger_{\mathscr{P}}$. The full-domain fallback when no cell is accepted cannot reduce coverage.

Sufficient endpoint limits. For a centered normal limit, $M(l_N) \le 0$ implies

$$\{L_M(l_N) > 0\} \subseteq \left\{ \frac{\sqrt{N}\{\hat{M}(l_N) - M(l_N)\}}{\hat{\sigma}_M(l_N)} > z_{1-\alpha/2} \right\},$$

and the right test has the analogous lower-tail inclusion. □

Lemma 3 (Boundary support consistency)**.** *Under Assumption 1(a)–(b), $\hat{A}_b$ converges uniformly to A_b on $[0,1]$, and $\hat{\sigma}_A(r_N) = O_p(1)$ along any deterministic sequence of shares in a compact subset of $(0,1)$.*

Proof. Uniform balance. Because $A_b(p) = -\int_0^{1-p} F^{-1}_{q_b}(s)\, ds$, the uniform estimation error is bounded by the integrated absolute quantile difference, the Wasserstein-1 distance W_1 between the fitted and population score laws:

$$\sup_{p\in[0,1]} |\hat{A}_b(p) - A_b(p)| \le W_1(\hat{P}^{\hat{q}}_N, P_{q_b}),$$

where $\hat{P}^{\hat{q}}_N$ is the empirical law of the fitted scores. Coupling fitted and oracle scores observation by observation costs at most $P_N|\hat{q} - q_b| = o_p(1)$ by nuisance consistency and the score envelope, and the oracle empirical score law converges in Wasserstein-1 by the first-moment empirical law; hence the left side is $o_p(1)$.

Cutoff bound. At an interior share r, the cutoff inequalities give

$$|\hat{\xi}_r| \min(r, 1-r) \le P_N |\hat{q}|,$$

so $\hat{\xi}_{r_N} = O_p(1)$ along any deterministic sequence in a compact subset of $(0,1)$.

Variance bound. The envelope, the bounded fractional weights, and empirical mean-square consistency of the fitted nuisance influences bound the empirical second moment of the fitted variance score by $O_p(1)$; hence $\hat{\sigma}_A(r_N) = O_p(1)$. □

Proof of Theorem 2. Set $M = A_b$ in Corollary 1; its sign conditions follow from Lemma 1 and Proposition 1. At a crossing in the reporting range, including either reporting boundary, the vanishing mesh puts every tested endpoint of the chosen cell near $p^\dagger$, and Proposition 2 gives its sequential studentized limit. For a crossing strictly outside the reporting range, the outer nodes eventually lie on the same side of the crossing. The cell at that end has one tested endpoint, which converges to the nearer reporting boundary as the mesh vanishes. Continuity and Lemma 1 give a strictly signed balance at that boundary. By Lemma 3, the estimated balance converges to that value and the test half-width vanishes at rate $O_p(N^{-1/2})$, so the probability of discarding the cell tends to zero. Thus the clipped hull covers $p^\dagger_{\mathscr{D}}$ with the stated asymptotic probability. On that event, $p \in (\underline{p}, \hat{p}_L)$ implies $p < p^\dagger$, and $p \in (\hat{p}_U, \overline{p})$ implies $p > p^\dagger$. Proposition 1 gives the preservation and overturning ranges; either range is empty if its corresponding interval endpoint lies on that side's reporting boundary. A measurable subset of the prescribed mass overturns the conclusion if it attains the tail bound in the overturning range. If the probability of reporting no interval tends to zero, counting those cases as noncoverage reduces the coverage probability by at most $o(1)$. □

A.5. Information restrictions and retained laws

Finally, we relate unrestricted fractional exclusions $\mathscr{D}_p$ to covariate-restricted rules and contaminated-sampling neighborhoods. If exclusion can use only

information V, the relevant score is conditional support $m_V := \mathbb{E}[q_b \mid V]$, and we have

$$\sup_{0\le d(V)\le 1,\ \mathbb{E}[d(V)]=p} \mathbb{E}[d(V)q_b] = \mathrm{T}_p(m_V) \le \mathrm{T}_p(q_b), \tag{11}$$

where T_p is the upper-tail integral defined in (2). In (11), the equality follows from iterated expectation and upper-tail allocation. The inequality follows because every $\sigma(V)$-measurable rule belongs to $\mathscr{D}_p$. An estimated ranking based only on baseline information cannot remove more population support at the same budget than ranking by true conditional support; even that oracle ranking is bounded by outcome-dependent selection. Richer information weakly lowers the primitive tipping fraction; covariate-measurable subpopulations (see Jeong and Namkoong, 2020) correspond to $V = X$. A literal subset attains the upper-tail bound if it can include all scores strictly above the cutoff and enough of the tied mass to reach p. Fractional weights attain the bound even when no such subset exists; when such a subset exists, it overturns the conclusion whenever the fractional optimum does.

The retained laws form $\mathscr{Q}_p := \{\mathbb{Q} : (1-p)\mathbb{Q} \le P\}$, where the inequality holds on every measurable set. Equivalently, their densities relative to P are bounded by $1/(1-p)$. This is the contaminated-sampling family of Horowitz and Manski (1995), used here to describe compositional loss among existing types. The support balance gives

$$\inf_{\mathbb{Q}\in\mathscr{Q}_p} \mathbb{E}_{\mathbb{Q}}[q_b] = \frac{\mathbb{E}[q_b] - \mathrm{T}_p(q_b)}{1-p} = -\frac{A_b(p)}{1-p}, \qquad 0 \le p < 1. \tag{12}$$

When the reference adjustments and floor are known, the retained-law problem reduces to a scalar trimmed-expectation problem in the framework of Palomba (2026).

Remark 2 (The deletion neighborhood)**.** *For an exclusion weight $d \in \mathscr{D}_p$ with $p \in (0,1)$, the retained and deleted laws P^{ret} and P^{del} have densities $(1-d)/(1-p)$ and d/p relative to P. Conversely, every retained law with density ratio $dP^{\mathrm{ret}}/dP \le 1/(1-p)$ arises from $d = 1-(1-p)dP^{\mathrm{ret}}/dP$ with $\mathbb{E}[d] = p$. The original law is the mixture $P = (1-p)P^{\mathrm{ret}} + pP^{\mathrm{del}}$.*

B. Sensitivity Curve: Characterization and Computation

We extend the reference-based profile to linear contrasts and use it to compute how far a contrast can fall at a given loss budget.

Let $\{(x_i, y_i)\}_{i=1}^N \subset \mathbb{R}^d \times \mathbb{R}$ be i.i.d. draws from P, with $\mathbb{E}[y^2] < \infty$, $\mathbb{E}[\|x\|^4] < \infty$, and $Q := \mathbb{E}[xx'] \succ 0$. Define

$$\beta := Q^{-1}\mathbb{E}[xy], \qquad u := y - x'\beta, \qquad \sigma^2(x) := \mathbb{E}[u^2 \mid x], \tag{13}$$

with $\mathbb{E}[\|x\|^2 \sigma^2(x)] < \infty$. Here $\sigma^2(x)$ is the conditional second moment of u, not necessarily its conditional variance: the projection need not satisfy $\mathbb{E}[u \mid x] = 0$. The normal equation

$$\mathbb{E}[xu] = 0 \tag{14}$$

defines the projection without a linear conditional mean assumption; nonlinear conditional means and heteroskedasticity are allowed. For a fixed nonzero vector $a \in \mathbb{R}^d$, write

$$\vartheta = a'\beta, \qquad r_a := Q^{-1}a, \qquad \mu_Z(a) := a'Q^{-1}a = \mathbb{E}[(r_a'x)^2].$$

A single coefficient corresponds to $a = e_1$, and a fitted value at x_0 corresponds to $a = x_0$. The profile sets the benchmark using the reference population.

Proposition 3 (Least-squares profile for a contrast)**.** *For $Q \succ 0$, $a \neq 0$ and every $t \in \mathbb{R}$, the unique constrained population least-squares solution is*

$$\beta(t) := \arg\min_{\gamma:\, a'\gamma = t} \mathbb{E}[(y - x'\gamma)^2] = \beta + \frac{t - a'\beta}{a'Q^{-1}a} Q^{-1}a.$$

The profile satisfies $\beta(\vartheta) = \beta$. Equivalently, the unique coefficient direction minimizing the mean-squared change in fitted values for a unit change in the contrast is $Q^{-1}a/(a'Q^{-1}a)$.

Proof of Proposition 3. Projection orthogonality gives $\mathbb{E}[(y - x'\gamma)^2] = \mathbb{E}[u^2] + (\gamma - \beta)'Q(\gamma - \beta)$. Setting $d_a = r_a/\mu_Z(a)$, any feasible vector γ satisfies $\gamma - \beta = (t - \vartheta)d_a + h_0$ with $a'h_0 = 0$, and its quadratic loss is $(t - \vartheta)^2/\mu_Z(a) + h_0'Qh_0$ because $d_a'Qh_0 = 0$. Positive definiteness makes $h_0 = 0$ the unique minimizer; applying the same decomposition to a unit contrast change $t - \vartheta = 1$ proves the second statement. □

The normalized influence and leverage variables for this profile are

$$Y(a) := (r_a'x)u, \qquad Z(a) := \frac{(r_a'x)^2}{\mu_Z(a)}, \qquad \mathbb{E}[Y(a)] = 0, \quad \mathbb{E}[Z(a)] = 1. \tag{15}$$

Along the profile, the residual at contrast t is $u - (t - \vartheta)r_a'x/\mu_Z(a)$. Minimizing its retained mean square therefore gives an adverse movement of $\mathbb{E}[Y(a)d]/\mathbb{E}[Z(a)(1-d)]$ whenever the denominator is positive, using $\mathbb{E}[Y(a)] = 0$. For the coefficient on w, with $x = (w, z')'$, the Frisch–Waugh–Lovell theorem gives $r_a'x = v/L$ and $\mu_Z(a) = 1/L$, recovering (7).

B.1. The Dinkelbach–threshold reformulation

Start with a fixed array $(Y_i, Z_i)_{i=1}^N$ with $Z_i \geq 0$ and a deletion count $1 \leq k < N$. For $\mathscr{S} := \{S \subset \{1, \dots, N\} : |S| = k\}$, write $Y_S = \sum_{i \in S} Y_i$ and $Z_{-S} = \sum_{i \notin S} Z_i$. The Dinkelbach transform replaces the ratio Y_S/Z_{-S} by an affine criterion and a modified score:

$$F_N(\lambda) := \max_{S \in \mathscr{S}} (Y_S - \lambda Z_{-S}), \qquad W_i(\lambda) := Y_i + \lambda Z_i.$$

The entries need not be independent or even random.

Proposition 4 (Fractional representation)**.** *Under the condition* $\mathscr{D}_N := \{Z_{-S} > 0 \; \forall \, S \in \mathscr{S}\}$*:*

(i) F_N *is finite, continuous, convex, and strictly decreasing;*

(ii) $\hat{\delta} = \max_{S \in \mathscr{S}} Y_S/Z_{-S}$ *is the unique root of* $F_N(\lambda) = 0$*;*

(iii) any maximizer $S^* \in \arg\max_S Y_S/Z_{-S}$ also solves $S^* \in \arg\max_S \sum_{i\in S} W_i(\hat{\delta})$, and hence consists of the k observations with the largest values of $W_i(\hat{\delta})$, with ties at the cutoff allocated arbitrarily (breaking ties by observation index selects one maximizer); conversely, every such top-k set attains the ratio $\hat{\delta}$.

Proof. On $\mathscr{D}_N$, F_N is the maximum of finitely many affine functions $\lambda \mapsto Y_S - \lambda Z_{-S}$ with strictly negative slopes $-Z_{-S} < 0$, so it is finite, continuous, convex, and strictly decreasing. At $\hat{\delta} = \max_{S\in\mathscr{S}} Y_S/Z_{-S}$, every subset satisfies $Y_S - \hat{\delta}Z_{-S} \le 0$ with equality if and only if $S \in \arg\max_{S'\in\mathscr{S}} Y_{S'}/Z_{-S'}$, giving $F_N(\hat{\delta}) = 0$ and uniqueness by strict monotonicity. Moreover, $Y_S - \hat{\delta}Z_{-S} = \sum_{i\in S} W_i(\hat{\delta}) - \hat{\delta}\sum_{i=1}^N Z_i$, so maximizing $Y_S - \hat{\delta}Z_{-S}$ over $S \in \mathscr{S}$ is equivalent to maximizing $\sum_{i\in S} W_i(\hat{\delta})$, whose maximizers are exactly the top-k sets. □

Lemma 4 (Top-k threshold representation)**.** *For reals $V_{[1]} \ge \cdots \ge V_{[N]}$ in descending order and $1 \le k \le N$,*

$$\sum_{j=1}^{k} V_{[j]} = \min_{\xi\in\mathbb{R}} \Big\{ \textstyle\sum_{i=1}^N (V_i - \xi)_+ + k\xi \Big\},$$

with minimizers $\{\xi : \#\{i : V_i > \xi\} \le k \le \#\{i : V_i \ge \xi\}\}$.

Applying Lemma 4 to $W_i(\lambda)$ turns $F_N(\lambda)$ into a sum of the top k scores minus $\lambda\sum_i Z_i$. Writing the normalized criterion as a threshold minimization, with $p_N = k/N$, gives

$$\begin{aligned} \Gamma_N(\lambda,\xi) &:= \frac{1}{N}\sum_{i=1}^{N}\big[(Y_i + \lambda Z_i - \xi)_+ + p_N\xi - \lambda Z_i\big], \\ f_N(\lambda) &:= \frac{F_N(\lambda)}{N} = \min_{\xi\in\mathbb{R}} \Gamma_N(\lambda,\xi). \end{aligned} \tag{16}$$

Fitted influence and leverage. On a sample with nonsingular $\hat{Q}_N := N^{-1}\sum_i x_i x_i'$, write $\hat{\beta} := \hat{Q}_N^{-1}N^{-1}\sum_i x_i y_i$, $\hat{u}_i := y_i - x_i'\hat{\beta}$, $\hat{r}_a := \hat{Q}_N^{-1}a$ and $\hat{\mu}_Z(a) := a'\hat{Q}_N^{-1}a$. The fitted influence score $\hat{s}_i(a) = \hat{r}_a' x_i \hat{u}_i$ and normalized leverage $\hat{Z}_i(a) = (\hat{r}_a' x_i)^2/\hat{\mu}_Z(a)$ satisfy $\sum_i \hat{s}_i = 0$ and $N^{-1}\sum_i \hat{Z}_i = 1$. At $k = \lfloor pN \rfloor$, the ratio for

these scores is

$$\hat{\delta}(p) := \max_{S \in \mathscr{S}} \frac{\sum_{i \in S} \hat{s}_i(a)}{\sum_{i \notin S} \hat{Z}_i(a)}. \tag{17}$$

Let $T_{N,p}(V) := N^{-1} \sum_{j=1}^{k} V_{[j]}$ and $h_Z := T_{N,p}(\hat{Z})$. Here $h_Z < 1$ certifies positive retained directional variation for every size-$(N-k)$ sample, a weaker requirement than nonsingularity of every retained full design.

B.2. Curve computation with a numerical error bound

On the domain $h_Z < 1$, Proposition 4 identifies $\hat{\delta}(p)$ with the unique root of $f_N(\lambda) = T_{N,p}(\hat{s} + \lambda \hat{Z}) - \lambda$. The score ranking depends on the candidate root; unlike scalar tipping, computing each curve value requires a new scalar solve. Fractional and integer tails agree at count nodes. Each evaluation sorts $\hat{s}_i + \lambda \hat{Z}_i$ and sums the largest k, so bisection on the bracket $[T_{N,p}(\hat{s}),\ T_{N,p}(\hat{s})/(1-h_Z)]$ of width $H = T_{N,p}(\hat{s}) h_Z/(1-h_Z)$ reaches absolute tolerance ε in $O(1+\log(1+H/\varepsilon))$ evaluations. Each candidate set retains at least $1 - h_Z$ normalized leverage, so the maximum of the affine criteria satisfies $f_N(\lambda_2) \le f_N(\lambda_1) - (1-h_Z)(\lambda_2 - \lambda_1)$ for $\lambda_2 > \lambda_1$. Thus $|f_N(\lambda)| \le (1-h_Z)\varepsilon$ certifies absolute root error at most ε; taking $\varepsilon = o(N^{-1/2})$ makes numerical error negligible for inference. The selected set is exactly optimal if it remains top-k at its own attained ratio (Proposition 4).

Lemma 5 (The empirical curve is nondecreasing)**.** *For $1 \le k < N-1$, if every size-$(N-k-1)$ retained sample has positive directional leverage, then $\hat{\delta}((k+1)/N) \ge \hat{\delta}(k/N)$.*

Proof. The leverage condition also holds at count k, since leverage is nonnegative, so both criteria have unique roots by Proposition 4. At the budget-k root, the selected modified-score sum divided by N equals the full modified-score mean, so the remaining $N-k$ retained scores sum to zero. Their largest score is therefore nonnegative. Adding it makes the budget-$(k+1)$ criterion nonnegative at that root; because the criterion is strictly decreasing, it has a weakly larger root. □

Rescaling for the derivative formulas. The fitted design derivative (Online Supplement, (S5)) uses an unnormalized ratio. With the contrast index $\hat{t}_{a,i} := \hat{r}_a' x_i$ (equal to $\hat{v}_i/\hat{L}$ for the coefficient on w), define

$$\hat{\delta}^{\mathrm{sc}} := \max_{S \in \mathscr{S}} \frac{\sum_{i \in S} \hat{t}_{a,i} \hat{u}_i}{\sum_{i \notin S} \hat{t}_{a,i}^2}. \tag{18}$$

The maximizers are unchanged, and no second solve is needed: $\hat{\mu}_Z(a) > 0$ is common to all subsets, so

$$\hat{\delta}(p) = \hat{\mu}_Z(a)\, \hat{\delta}^{\mathrm{sc}}. \tag{19}$$

B.3. Population targets

For a fixed share $p \in (0,1)$, write $\tau = 1 - p$ and $k = \lfloor pN \rfloor$. The quantile convention is that of (2). Sampling assumptions enter when we ask whether the array criterion is well defined with high probability and how its root relates to the population root.

Assumption 2 (Data-generating process)**.** *(Y_i, Z_i) are i.i.d. with common distribution (Y, Z); $Z_i \geq 0$ a.s.; and $\mathbb{E}[Y^2] < \infty$, $\mathbb{E}[Z^2] < \infty$.*

For the normalized directional pair in (15), $m_Z := \mathbb{E}[Z] = 1$; for a general pair (Y, Z), write $m_Z = \mathbb{E}[Z]$.

Assumption 3 (Denominator non-degeneracy)**.** *With $m_Z := \mathbb{E}[Z]$, assume $m_Z > 0$ and $F_Z^{-1}(\tau) > 0$. For $Z \geq 0$, the condition $F_Z^{-1}(\tau) > 0$ is equivalent to $\int_0^\tau F_Z^{-1}(u)\, du > 0$ and to $P(Z = 0) < \tau$, and itself implies $m_Z > 0$.*

Under Assumptions 2 and 3, the smallest retained denominator divided by N is $N^{-1} \sum_{j=k+1}^{N} Z_{[j]}$, which converges almost surely to $\int_0^\tau F_Z^{-1}(u)\, du > 0$, so $\mathbb{P}(\mathscr{D}_N) \to 1$.

Assumption 4 (Positive upper-tail numerator)**.** $\mu_p^Y := \int_\tau^1 F_Y^{-1}(u)\, du > 0$.

For the directional pair $(Y(a), Z(a))$, Assumption 3 holds exactly when $P(r_a' x = 0) < \tau$, and Assumption 4 follows from mean zero and nondegeneracy by Lemma 1.

In the population dual (2), the minimizing cutoffs for a score V satisfy $P(V > \xi) \le p \le P(V \ge \xi)$, and $\mathrm{T}_p(V)/p$ is the upper-tail mean. With $W(\lambda) = Y + \lambda Z$, applying this duality to the population counterpart of (16) gives

$$f(\lambda) := \min_{\xi \in \mathbb{R}} \mathbb{E}[(W(\lambda) - \xi)_+ + p\xi - \lambda Z] = \mathrm{T}_p(W(\lambda)) - \lambda m_Z. \tag{20}$$

Proposition 5 (Existence and uniqueness of the population root)**.** *For an integrable pair (Y, Z) with $Z \ge 0$ satisfying Assumption 3, $f(\lambda) = \mathrm{T}_p(Y + \lambda Z) - \lambda m_Z$ is continuous and strictly decreasing on $\mathbb{R}$, with $\lim_{\lambda \to -\infty} f(\lambda) = +\infty$ and $\lim_{\lambda \to \infty} f(\lambda) = -\infty$. Hence f has a unique real root δ, which is strictly positive if and only if Assumption 4 holds; for the nondegenerate directional pair, positivity holds automatically (Lemma 1). This root equals the fractional ratio value*

$$\delta = \sup_{0 \le d \le 1,\ \mathbb{E}[d] = p} \frac{\mathbb{E}[Yd]}{\mathbb{E}[Z(1-d)]}.$$

If at the root there is a cutoff ξ with $P(W(\delta) \ge \xi) = p$, then the supremum is attained by the region $D = \{W(\delta) \ge \xi\}$ and coincides with the supremum over measurable regions of mass p. For $(Y(a), Z(a))$ with $m_Z = 1$, the fractional ratio value is δ.

Proof. By Assumption 3, $c_Z := m_Z - \mathrm{T}_p(Z) = \int_0^\tau F_Z^{-1}(u)\, du > 0$. For any $\lambda_1 < \lambda_2$, subadditivity and positive homogeneity of T_p give

$$\mathrm{T}_p(Y + \lambda_2 Z) = \mathrm{T}_p\big(Y + \lambda_1 Z + (\lambda_2 - \lambda_1)Z\big) \le \mathrm{T}_p(Y + \lambda_1 Z) + (\lambda_2 - \lambda_1)\mathrm{T}_p(Z),$$

so subtracting $\lambda_2 m_Z$ yields $f(\lambda_2) \le f(\lambda_1) - (\lambda_2 - \lambda_1)c_Z$. Combined with the Lipschitz bound $|f(\lambda_1) - f(\lambda_2)| \le 2m_Z|\lambda_1 - \lambda_2|$ (since $|\mathrm{T}_p(V) - \mathrm{T}_p(B)| \le \mathbb{E}[|V - B|]$ and $Z \ge 0$), this strict descent gives $\lim_{\lambda \to -\infty} f(\lambda) = +\infty$, $\lim_{\lambda \to \infty} f(\lambda) = -\infty$, and a unique real root δ, which is strictly positive if and only if $f(0) =$

$\mathrm{T}_p(Y) > 0$. For any feasible weight $d \in [0,1]$ with $\mathbb{E}[d] = p$, upper-tail duality gives $\mathbb{E}[Z(1-d)] = m_Z - \mathbb{E}[Zd] \geq m_Z - \mathrm{T}_p(Z) = c_Z > 0$, and

$$\mathbb{E}[Yd] - \delta\mathbb{E}[Z(1-d)] = \mathbb{E}[W(\delta)d] - \delta m_Z \leq \mathrm{T}_p(W(\delta)) - \delta m_Z = f(\delta) = 0,$$

so $\mathbb{E}[Yd]/\mathbb{E}[Z(1-d)] \leq \delta$. Choosing the upper-tail weight $d_p^* = \mathbf{1}\{W(\delta) > \xi\} + \alpha\mathbf{1}\{W(\delta) = \xi\}$, where ξ is a $(1-p)$-quantile of $W(\delta)$ and $\alpha \in [0,1]$ satisfies $\mathbb{E}[d_p^*] = p$, achieves equality $\mathbb{E}[W(\delta)d_p^*] = \mathrm{T}_p(W(\delta))$, so its ratio attains δ. When $P(W(\delta) \geq \xi) = p$, we may take $\alpha = 1$, giving $d_p^* = \mathbf{1}\{W(\delta) \geq \xi\}$ and establishing region attainment. □

The following assumption ensures attainment by an upper-tail region of mass p at every nonnegative multiplier, including the directional root.

Assumption 5 (Population threshold regularity)**.** *For every $\lambda \geq 0$, $W(\lambda) = Y + \lambda Z$ has a unique minimizing cutoff $\xi(\lambda) = F^{-1}_{W(\lambda)}(\tau)$ in (20), and $P(W(\lambda) \geq \xi(\lambda)) = p$. (On compact λ-sets, a density that is positive and continuous near $F^{-1}_{W(\lambda)}(\tau)$ uniformly over the set is sufficient.)*

The curve central limit theorem requires cutoff regularity only near the population root (Online Supplement, Assumption S1). All top-p allocations of the root score attain the same ratio, though they can leave different identifying variation when Z varies among observations tied at the cutoff. Empirical ties alone do not establish whether the population has a cutoff atom.

B.4. One criterion for movement and tipping

For the normalized pair with $\mathbb{E}[Z] = 1$, write

$$F(\lambda, p) := \mathrm{T}_p(Y + \lambda Z) - \lambda. \tag{21}$$

On the positive-leverage domain, fix p and solve $F(\lambda, p) = 0$ for movement $\lambda = \delta(p)$; under Proposition 1, fix the margin $\lambda = c$ and find the first zero in p for tipping. For the coefficient pair in (7), $q_b = L(Y + cZ)$ and $\mathbb{E}[q_b] = Lc$, so

positive homogeneity gives

$$A_b(p) = \mathrm{T}_p(q_b) - \mathbb{E}[q_b] = LF(c,p). \tag{22}$$

Proposition 6 (Curve and support identify the same crossing)**.** *For the coefficient pair in (7), let* $p \in \mathscr{P}_+ := \{p \in (0,1) : \mathrm{T}_p(Z) < 1\}$*,* $c = \vartheta - b > 0$*, and* $\mathbb{E}[|Y|] < \infty$*. Then*

$$\delta(p) \geq c \quad \Longleftrightarrow \quad \mathrm{T}_p(Y + cZ) \geq c \quad \Longleftrightarrow \quad A_b(p) \geq 0. \tag{23}$$

Under Proposition 1, $p^\dagger \in \mathscr{P}_+$*,* $\delta(p^\dagger) = c$*, and* $p^\dagger = \inf\{p \in \mathscr{P}_+ : \delta(p) \geq c\}$*.*

Proof. On $\mathscr{P}_+$, $F(\cdot, p)$ decreases strictly through its unique root by Proposition 5, so $\delta(p) - c$ has the sign of $F(c,p)$ and hence of $A_b(p)$ by (22). Since $Z \geq 0$ and $\mathbb{E}[Z] = 1$, $\mathscr{P}_+ = (0, P(Z > 0))$. The inequalities $p^\dagger < P(q_b > 0) \leq P(Z > 0)$ place $p^\dagger$ in this domain, where the first zero of A_b yields the claimed curve crossing. □

On the coverage event, the tipping interval also gives sign statements about $\delta(p) - c$ on $\mathscr{P}_+$, not simultaneous coverage of curve magnitudes.

For the fitted pair in (17) and $0 < p < 1$, define the fractional empirical criterion

$$\hat{F}_N(\lambda, p) := \min_{\xi \in \mathbb{R}} \left\{ \frac{1}{N} \sum_i (\hat{s}_i + \lambda \hat{Z}_i - \xi)_+ + p\xi \right\} - \lambda. \tag{24}$$

For the coefficient on w, $\hat{q}_i = \hat{L}(\hat{s}_i + \hat{c}\hat{Z}_i)$ and $N^{-1}\sum_i \hat{q}_i = \hat{L}\hat{c}$, so (8) gives $\hat{A}_b(p) = \hat{L}\hat{F}_N(\hat{c}, p)$ throughout $(0,1)$ when both calculations use the same orientation and fitted floor. At $p = k/N$ with $1 \leq k < N$, $\hat{F}_N$ equals the count criterion f_N in (16), so the sign of $\hat{A}_b(p)$ agrees with that of $\hat{\delta}(p) - \hat{c}$ on the positive-leverage domain. Between count nodes, $\hat{F}_N$ adds $(p - k/N)$ times the $(k+1)$-st largest modified score, for $k = \lfloor Np \rfloor$, so the sign of $\hat{A}_b(p)$ need not agree with that of $\hat{\delta}(p) - \hat{c}$.

B.5. AMIP bounds and reranking

Proposition 7 (Relation to the first-order approximation)**.** *For the directional pair, assume* $\mathbb{E}[Z(a)] = 1$ *and let* $\delta \in [0, \infty)$ *satisfy* $\mathrm{T}_p(Y(a) + \delta Z(a)) = \delta$*. If* $\mathrm{T}_p(Z(a)) < 1$*, then*

$$\mathrm{T}_p(Y(a)) \leq \delta \leq \frac{\mathrm{T}_p(Y(a))}{1 - \mathrm{T}_p(Z(a))}.$$

The lower bound is the first-order AMIP approximation; its sample analogue is $N^{-1} \sum_{j=1}^{k} \hat{s}_{[j]}(a)$*, the normalized sum of the* k *largest fitted influence scores. The same inequalities hold for the fitted finite-sample pair.*

Proof of Proposition 7. Since $Z(a) \geq 0$ and $\delta \geq 0$, monotonicity of T_p gives $\delta = \mathrm{T}_p(Y + \delta Z) \geq \mathrm{T}_p(Y)$. Subadditivity and positive homogeneity give $\delta \leq \mathrm{T}_p(Y) + \delta \mathrm{T}_p(Z)$, yielding the upper bound by rearrangement. The same argument at the empirical root uses monotonicity, subadditivity, and homogeneity of $T_{N,p}$, which follow from its maximum over subsets of size k. □

The two bounds reflect different observation rankings, so neither need be tight. Let d_p^* be an attaining exclusion weight at budget p (Proposition 5), $\nu^* := \mathbb{E}[Z(a) d_p^*]$ its deleted leverage, and $B_a(p) = \mathbb{E}[Z(a)(1 - d_p^*)] = 1 - \nu^*$ its retained leverage. The root identity $\mathbb{E}[\{Y(a) + \delta Z(a)\} d_p^*] = \delta$ rearranges to $\mathbb{E}[Y(a) d_p^*] = \delta B_a(p)$, yielding the exact decomposition

$$\delta = \frac{\mathrm{T}_p(Y(a))}{B_a(p)} - \frac{\mathrm{T}_p(Y(a)) - \mathbb{E}[Y(a) d_p^*]}{B_a(p)}. \tag{25}$$

Here the subtracted term is a nonnegative reranking loss divided by retained leverage: because d_p^* is a feasible weight of mass p, $\mathrm{T}_p(Y(a))$ maximizes deleted influence over all such weights, and the loss vanishes if and only if d_p^* is also a top-p weight of $Y(a)$. Since $\nu^* \leq \mathrm{T}_p(Z(a))$ and the reranking loss is nonnegative, (25) refines the upper bound of Proposition 7 via the tighter bound $\mathrm{T}_p(Y(a))/B_a(p)$. The upper bound of Proposition 7 is attained when d_p^* is simultaneously a top-p weight of $Y(a)$ and of $Z(a)$; for constant leverage $Z(a) \equiv 1$, one has $\nu^* = p$ and $\delta = \mathrm{T}_p(Y(a))/(1 - p)$ exactly. If

$P(W^* > \xi) < p < P(W^* \geq \xi)$ for the root score $W^* = Y(a) + \delta Z(a)$ and cutoff ξ, the rule allocates only part of the cutoff cell, and ν^* uses the leverage of that fractional allocation.

Whenever $\mathrm{T}_p(Y(a)) > 0$ (so both bounds are positive), their ratio $1/\{1 - \mathrm{T}_p(Z(a))\}$ tends to 1 as $p \to 0$, since $\mathrm{T}_p(Z(a)) \to 0$.

C. Unrestricted-Refitting Comparisons

Refitting lets the retained observations determine the control adjustment. How much the focal coefficient changes depends on the relationships the controls have with the adjusted regressor and the residual after exclusion.

Proposition 8 (Refitting difference)**.** *Let $\mathbb{E}_R$ denote expectation under the retained law $R\,dP/\mathbb{E}[R]$ for a retention weight $R \in [0,1]$ with $\mathbb{E}[R] > 0$. Suppose the retained full design is nonsingular, and let $\vartheta_R^{\mathrm{full}}$ denote its unrestricted refitted coefficient on w. Define*

$$a_R = \mathbb{E}_R[zv], \qquad m_{zu,R} = \mathbb{E}_R[z\{h - \vartheta(R)v\}].$$

Write $H_R = \mathbb{E}_R[zz']$, $L_R = \mathbb{E}_R[v^2]$ and $S_R = L_R - a_R' H_R^{-1} a_R$ for the retained control moment matrix, identifying variation and variation after re-adjusting the regressor, respectively. Then $H_R \succ 0$ and $S_R > 0$, and

$$\vartheta_R^{\mathrm{full}} - \vartheta(R) = -\frac{a_R' H_R^{-1} m_{zu,R}}{S_R}. \tag{26}$$

Proof. Replacing (w, y) by $(v, h) = (w - z'\pi, y - z'\eta)$ leaves the coefficient on v in the unrestricted regression on (v, z) equal to $\vartheta_R^{\mathrm{full}}$. The change from (w, z) to (v, z) is invertible, so the retained moment matrix remains positive definite, with control block $H_R \succ 0$ and Schur complement $S_R > 0$. The Frisch–Waugh–Lovell theorem then gives

$$\vartheta_R^{\mathrm{full}} = \frac{\mathbb{E}_R[vh] - a_R' H_R^{-1} \mathbb{E}_R[zh]}{S_R}, \qquad \mathbb{E}_R[zh] = m_{zu,R} + \vartheta(R) a_R.$$

Substituting $\mathbb{E}_R[vh] = \vartheta(R)L_R$ and the expression for $\mathbb{E}_R[zh]$ yields (26). □

The identity also holds for empirical moments and fractional retention weights. If the adjusted regressor stays orthogonal to the controls after exclusion ($a_R = 0$), the two coefficients agree even when the outcome adjustment changes. The numerator has no sign restriction, and a small numerator need not imply a small gap when S_R is small. Because (26) compares retained coefficients, differences in adverse movements $\vartheta - \vartheta_R^{\text{full}}$ versus $\vartheta - \vartheta(R)$ take the opposite sign. For a general contrast, an invertible change of coefficient coordinates with first coordinate $a'\beta$ reduces the comparison to this coefficient case.

Remark 3 (Agreement with refitting in randomized comparisons)**.** *Let $w = D$ be a binary treatment with $D \perp\!\!\!\perp (Y(1), Y(0), X)$ and $P(D = 1) = e \in (0, 1)$. Let $y = DY(1) + (1 - D)Y(0)$ and let z consist of an intercept and functions of covariates X. If $R = s(X)$, then $v = D - e$ and $\mathbb{E}[R\, z v] = 0$. Under a nonsingular retained design, both the coefficient with fixed adjustments and the unrestricted refit equal $\mathbb{E}[R\,\tau(X)]/\mathbb{E}[R]$, where $\tau(X) = \mathbb{E}[Y(1) - Y(0) \mid X]$. Outcome- or treatment-dependent loss need not preserve this agreement.*

D. Primary Empirical Specifications

For Mindspark, the population interval maintains the finite-type observed-case model (Online Supplement, Section S1). STAR and ruggedness–income are evaluated as deterministic sample audits. Selected-set unrestricted refits are attained coefficients on the samples selected using fixed reference adjustments, not global optima over all subsets of the same size.

D.1. Mindspark: frame, scoring and population

The lottery assigned 619 students within 19 center–time-slot preference strata at three centers (Muralidharan, Singh, and Ganimian, 2019). The math regression uses 535 students with complete released baseline and endline

MLE scores: 264 offered students and 271 controls. Endline participation is 84.7% among offered students and 89.5% among controls. We regress endline achievement on an intercept, lottery offer and baseline achievement. The no-strata ANCOVA matches the offer coefficients and standard errors in the first two columns of Table 2 in the accepted manuscript of Muralidharan, Singh, and Ganimian (2019): 0.37 (0.064) for math and 0.23 (0.062) for Hindi, despite those columns' fixed-effect labels. With the released item-response-theory (IRT) scoring map and normalization fixed, finitely many response patterns give finitely many baseline and endline scores. Together with binary lottery offer, this implies finitely many record types. We maintain the singleton-cutoff, interior-split and nondegeneracy conditions of Proposition S1, not a continuous-score approximation. The sampling model has independent potential records within strata, fixed lottery allocations and student-specific response, targeting the design-weighted observed-case mixture (Corollary S1). The smallest stratum-by-offer cell has three recruits. The variance comparison is asymptotic, and we retain the i.i.d. standard error. The analysis conditions on usable scores and treats the released IRT scoring map as fixed, without modeling item-parameter uncertainty or missing outcomes. The supporting Hindi outcome uses the same specification with complete same-subject scores, giving $N = 537$. The Online Supplement reports equal-budget composition comparisons in Section S6.

D.2. STAR: analysis frame and retained variation

The women's fall-grade specification in Table 5 of Angrist, Lang, and Oreopoulos (2009) includes an intercept, the three offer indicators, and categorical controls for mother tongue, high-school group, November course count, the last-minute studying measure, and both parents' education. Each enters through indicators omitting its lowest observed coded level, giving 32 design columns. The chosen analysis frame has 729 women: 441 controls, 105 support-only, 121 incentive-only and 62 combined-offer women. It applies the released no-show exclusion and requires an observed fall grade, first-year GPA

and complete controls; fall-grade availability depends on taking one-semester courses. Because the controls include November course load, the resulting observed-case projection is not a baseline-defined subgroup treatment effect. Recorded grades and categorical controls induce an atomic support-score law; STAR remains a deterministic sample audit; no population cutoff-stability or sampling conditions are verified here.

For the half-effect set reported in Section 4.2, the selected-set refit is 1.8669, below the half-effect floor 2.1026, with 88.8% of the reference identifying variation retained. The half-effect and sign-crossing complements both retain rank 32.

Sparse controls and retained variation. Exactly one woman occupies the base course-count category. The coefficient vector γ_0 for the intercept minus the four included course indicators satisfies $X\gamma_0 = e_j$ for that row j and $a'\gamma_0 = 0$ for the combined-offer contrast. Deleting j lowers rank by one, to 31, leaving null space $\operatorname{span}\{\gamma_0\}$. Every deletion set containing j therefore leaves the unrestricted coefficient vector unidentified. The focal contrast nevertheless remains estimable after deleting j alone because $a'\gamma_0 = 0$, and

$$(X\hat{Q}_N^{-1}a)_j = N\gamma_0'a = 0$$

shows that row j contributes no identifying variation. The minimum retained fraction of identifying variation across all subsets remains above 0.09067 through 145/729 deletions, despite singular full-design complements. The analogous population distinction is between a zero-index atom and an atom at the active cutoff (Online Supplement, Proposition S5). Cutoff regularity for STAR's recorded grades remains unverified.

D.3. Ruggedness: finite-share comparison

For Nunn and Puga (2012), log GDP per capita is regressed on ruggedness, an Africa indicator and their interaction, controlling for diamonds, soil quality, tropical area, and coastal distance, each interacted with Africa. The Africa

differential slope is 0.32054 among 170 complete-case countries; the audit uses $b = 0$ as its floor (giving an initial margin of 0.32054), rather than a statistical significance threshold. Refitting the full regression after deleting the 10 countries selected with fixed reference adjustments yields a coefficient of -0.0684 on the interaction, which is an attained refit rather than a global optimum over all 10-country subsets. Through 51/170 deletions, all evaluated budgets pass the all-subset positive-denominator check and selected refits are nonsingular.

E. Calibration Methods

The numerical experiments evaluate coverage and how often no interval is reported under specified sampling laws, rather than testing the validity of a population model for either application. All primary scalar intervals use the direct-support procedure in Sections A and A.4, without multiplier simulation or a fitted-cutoff Wald interval. The Online Supplement records exact law constructions and numerical guards and reports the supplementary experiments (Section S5).

E.1. Primary scalar experiment

The primary experiment comprises 22 law–sample-size cells with 12,000 independent draws per cell. Mindspark math ($N = 535$) and Hindi ($N = 537$) each supply five laws, and STAR ($N = 729$) supplies six; evaluating each of these 16 application-based laws at both sign and half-effect floors gives 32 evaluations. Two synthetic laws at $N \in \{200, 800, 3200\}$, each evaluated at a single fixed floor, give another six evaluations, for 38 total evaluations. The reporting ranges are $[0.005, 0.20]$ for Mindspark, $[0.005, 0.10]$ for STAR, and $[0.02, 0.20]$ for the synthetic laws.

The Mindspark and STAR laws draw design rows from their analysis samples and vary the conditional outcome distribution, with one STAR law also altering the design weights (Section S5). The rounded and discrete-residual laws

serve as finite-support stress tests; their population targets allow fractional splitting of cutoff mass and do not validate the atomless-crossing inference theory for atomic populations. Continuous conditional outcomes can coexist with an atom at zero in the score distribution; similarly, these laws do not establish cutoff regularity for recorded STAR grades. The synthetic populations set $x = (1, T, C)'$ and $a = (0, 1, 0)'$, ordering the four covariate cells as $(T, C) = (0, 0), (0, 1), (1, 0), (1, 1)$. Their cell probabilities, conditional means, and conditional scales are, respectively,

$$(0.30, 0.35, 0.10, 0.25), \qquad (0, 0.2, 0.9, 2.3), \qquad (0.5, 1.2, 0.8, 1.5).$$

Outcomes equal the cell mean plus the cell scale times independent standard-normal or unstandardized Student-t_3 noise. The audited conclusion is that the projection contrast exceeds the fixed floor $b = 1$.

Targets and interval computation. Population curves and crossings are computed from conditional tail masses and first moments, rather than from simulated super-samples or from interpolating finite-sample count crossings. For continuous mixture components, we solve the cutoff-mass and modified-score tail equations; for discrete components, we sort the exact scores, retain their probability weights and split cutoff mass fractionally. Solver tolerances and numerical guards are recorded in Section S5.

Each sample is evaluated on padded count nodes k/N, running from $k = \lfloor N\underline{p} \rfloor$ through $k = \lceil N\overline{p} \rceil$. At each node we compute the fractional support balance, the complete fitted influence in (9), and its centered empirical second moment with divisor N. We then apply the grid-cell inversion rule of Section A.4 with critical value $z_{0.975}$. Coverage is defined as the resulting interval containing the clipped population crossing $p^{\dagger}_{\mathscr{P}}$. The procedure reports no interval if the estimated coefficient does not exceed the floor used in the audit, the fit fails, a standard error is nonpositive, or retained leverage is inadequate. Replications without an interval count as noncoverage. For the synthetic laws,

positive orientation requires a fitted contrast above 1; application-based laws require a positive fitted contrast prior to sign and half-effect evaluation.

In the primary experiment, 567 of 264,000 samples yielded no interval, all because they failed the orientation check. This occurred in 1 of 60,000 Mindspark math replications, 13 of 60,000 Mindspark Hindi replications, 377 of 72,000 STAR replications and 176 of 72,000 synthetic replications. In STAR, these failures concentrate in the two nonlinear heteroskedastic laws, with rates of 1.23% and 1.48%; 174 of the synthetic failures are in the Student-t_3 law at $N = 200$. A sample without an interval counts against both conclusions in the application-based laws and against one in the synthetic laws, giving 958 conclusion evaluations without an interval. The simultaneous exact binomial intervals allocate 0.05/76 to each tail across the 38 evaluations and are judged against the 0.94 lower-bound benchmark. A separate small-N Student-t_3 confirmation with 100,000 replications uses 0.025 in each binomial tail and is not pooled into the original simultaneous family (Online Supplement, Section S5.2).

Supplementary experiments. The Online Supplement reports tipping-interval stress tests for boundary crossings and weak orientation (Section S5.2), as well as estimated-margin curve calibration and a dependence diagnostic (Section S5.3).

F. Finite-Sample Limits and Cutoff Atoms

The pointwise intervals are asymptotic. The following construction shows that no procedure can combine uniform finite-sample coverage over the unrestricted finite-variance class with a width that is uniformly informative, even for the intercept-only projection with conclusion floor $b = 0$ and lower reporting boundary $\underline{p} = 0$.

Proposition 9 (No uniform finite-sample localization)**.** *Consider the intercept-only projection with conclusion floor $b = 0$ and reporting boundaries $\underline{p} = 0$ and $\overline{p} = 0.20$. For any interval procedure and any sample size N, there exist two*

distributions with finite variance and positive mean whose N-observation product laws are arbitrarily close in total variation, while their tipping fractions clipped to the reporting range differ by at least a positive constant independent of N. Consequently, for $\alpha < 1/2$, any procedure achieving coverage at least $1-\alpha$ under both distributions must have expected width bounded below by a positive constant independent of N.

Proof. Let P be uniform on $[1/2, 3/2]$, with mean one. The support scores are all positive, so no admissible exclusion reverses the mean, giving $p^\dagger = +\infty$ and clipped tipping fraction $\overline{p} = 0.20$. This law lies outside the negative-support condition in Assumption 1, as permitted by the unrestricted finite-variance class. For $\varepsilon > 0$ and $\mu = 0.01$, set

$$P_\varepsilon = (1-\varepsilon)P + \varepsilon\,\delta_{-M_\varepsilon}, \qquad M_\varepsilon = \frac{1-\varepsilon-\mu}{\varepsilon},$$

so that $\mathbb{E}_{P_\varepsilon}[y] = \mu > 0$ and the variance is finite for every fixed $\varepsilon > 0$. The first reversal removes positive values from the upper tail and, by Definition 1, equates deleted support to the full net support μ (not $\mu(1-p)$):

$$p\left(\frac{3}{2} - \frac{p}{2(1-\varepsilon)}\right) = \mu.$$

The small root tends to $p_0 = 3/2 - \sqrt{9/4 - 2\mu} \approx 0.00668$ as $\varepsilon \downarrow 0$. Any fixed upper reporting boundary above p_0 gives separation. The total-variation distance between the N-observation experiments is at most $N\varepsilon$. If the procedure has coverage at least $1-\alpha$ under each law, maximally couple the samples to agree with probability at least $1 - N\varepsilon$. Sample agreement and both coverage events occur together with probability at least $1 - 2\alpha - N\varepsilon$, on which the identical interval contains both separated tipping fractions. For each N, choose $\varepsilon = \varepsilon_N > 0$ with $N\varepsilon_N \to 0$. The experiments are then asymptotically indistinguishable while the clipped tipping fractions remain separated, so the lower bound on expected width is $(0.20 - p_0 - o(1))(1 - 2\alpha - o(1))$, with a positive limit as N grows. For sufficiently small ε, replacing δ_{-M_ε} with a smooth

density symmetric about $-M_\varepsilon$ and supported on $[-M_\varepsilon - 1, -M_\varepsilon + 1]$ preserves the mean, the upper-tail crossing, and the total-variation bound. □

This construction limits the type of finite-sample guarantee achievable, rather than contradicting the pointwise asymptotics: a known outcome or support bound, or a uniform tail envelope, rules out such adversarial point masses, whereas an empirical bound estimated solely from the observed range does not. We therefore report asymptotic pointwise guarantees and simulation evidence under the specified sampling laws, rather than a distribution-free finite-sample guarantee.

Cutoff atoms and bootstrap limits. The atomless scalar route requires a unique minimizing cutoff and positive complete influence variance (Assumption 1), but not a positive score density or a cutoff-estimation rate. The finite-type route permits a singleton cutoff type whose mass is split strictly inside it (Online Supplement, Proposition S1). Pointwise ratio inference and uniform sensitivity-curve bands require stronger cutoff regularity conditions (Online Supplement, Assumptions S1 and S3, respectively). A cutoff atom whose observations have different nuisance-score derivatives can induce a nonlinear first-order asymptotic distribution; standard Gaussian and nonparametric bootstrap inference can then fail, and adding random jitter alters the population target. Beyond the independent fixed-count stratified model in Corollary S1, dependent sampling and uncertainty from estimated item-response scores require further inference arguments; changing the bootstrap or enlarging the critical value does not by itself establish validity.