The Power Cost of Honest Confidence Intervals

A short guide to *Local Asymptotic Power of Honest Confidence Intervals*

Hugo Freeman

This note is a self-contained, informal summary of the paper's main definitions and results, built around easy examples. No proofs; those are in the paper. References appear inline.

1 Honest intervals, and the question

In many workhorse designs the estimator carries a bias that the data cannot reveal, and the standard fix is to *widen the interval by the worst case.* The analyst asserts an a priori bound: parallel trends can fail by at most M (difference-in-differences; Rambachan and Roth, *Review of Economic Studies*, 2023), the instrument can enter the outcome equation with a coefficient of at most $\bar{\gamma}$ (plausibly exogenous IV; Conley, Hansen and Rossi, *Review of Economics and Statistics*, 2012), the second derivative of the conditional mean is at most M near the cutoff (regression discontinuity; Armstrong and Kolesár, *Econometrica*, 2018), or at most R factors are weak (Armstrong, Weidner and Zeleneev, 2022). Writing $\bar{B}$ for the implied worst-case bias of the estimator, the resulting *honest* (bias-aware) confidence interval takes the form

$$\hat{\beta} \ \pm\ \big(\bar{B} + z_{1-\alpha/2} \cdot se(\hat{\beta})\big),$$

possibly with a sharpened critical value. Such intervals *cover*: they contain the true β with probability at least $1-\alpha$ no matter which admissible bias the world realised. That guarantee is their point, and this paper takes it as given.

The question the paper asks is the other side of the ledger: **what do these intervals cost in power?** Coverage is about keeping the truth inside; power is about expelling false values. An interval that never rejects anything has perfect coverage. Where between these poles do honest intervals sit?

The one-sentence answer: *every honest procedure carries a "dead zone" of alternatives, extending to twice the bias bound from the truth, inside which its worst-case power never exceeds the size α; whether this matters asymptotically is decided by a race between the bound and the standard error.*

2 Two numbers: the budget spent twice

The factor of two is the heart of the paper, and two numbers exhibit it. Set the bias bound to $\bar{B} = 1$ and ignore sampling noise.

Bias at the bound. Suppose the realised bias is also 1, so the estimate sits at $\hat{\beta} = \beta_0 + 1$. The honest interval is

$$\hat{\beta} \pm 1 \;=\; [\beta_0,\; \beta_0 + 2].$$

The truth is covered *exactly at the edge* (that is what a binding worst case looks like), and every value up to *twice* the bound above the truth is retained. Why must the interval reach so far? Because its far half defends not the world that generated the data but that world's observationally identical twin: a world with bias -1, in which the same data would have been produced by a truth at $\beta_0 + 2$. The data cannot distinguish the two worlds, so coverage in both is compulsory. Here the width is money well spent: with bias genuinely at the bound, coverage needs all of it.

No bias. Now suppose the realised bias is zero: the instrument is in fact valid, the trend truly parallel. The estimate sits at $\hat{\beta} = \beta_0$, yet the honest interval $[\beta_0 - 1, \beta_0 + 1]$ still retains every value within one bound of the truth, where an oracle who knew the bias were zero would spend only the sampling width $2z_{1-\alpha/2}\, se(\hat{\beta})$. And because the bias is untestable, this saving can never be identified or harvested.

So honesty prices the worst admissible world *in every world.* The dead zone is that price expressed in power rather than length: every retained value is a value the dual test cannot reject.

3 The dead zone

Setup in three rates. The estimator satisfies $\hat{\beta} - \beta = \hat{B} +$ noise, where $\hat{B}$ is the (unknowable) realised bias, the analyst's bound is $\bar{B} = \sup_\theta |\hat{B}|$, and se is the standard error. As n grows, se shrinks like $n^{-\epsilon}$ (parametric case: $\epsilon = \frac{1}{2}$) and $\bar{B}$ shrinks like $n^{-\delta}$ (possibly $\delta = 0$: a fixed bound). Power is measured against *local alternatives*, rival values of β approaching the truth at the scale the data can resolve, exactly as in textbook local asymptotics.

Definition (dead zone). The dead zone is the set of rival values that the honest test rejects no more often than its size α, *in the worst case over the admissible biases.* The paper's headline: the dead zone extends to $2\bar{B}$, twice the bound, on either side of the truth.

Per-law windows versus the union. The "twice" deserves care, because any single simulation or dataset displays only half of it. Under one particular world with realised bias $c\bar{B}$ (for some $c \in [-1, 1]$), the interval is centred at the estimate, which sits at $\beta_0 + c\bar{B}$, and it shelters values within $\bar{B}$ of *that centre*: a window of radius $\bar{B}$ whose location slides with the bias,

$$\big[(c-1)\bar{B},\ (c+1)\bar{B}\big] \qquad \text{(measured from the truth).}$$

Its far edge reaches $2\bar{B}$ when the bias is maximal. The $\pm 2\bar{B}$ dead zone is the union of these sliding windows over the admissible biases: the *guarantee-free* region. For any rival value inside it, some admissible bias hides that rival completely; and since the bias is untestable, only the union can be promised. Figure 1 shows both objects.

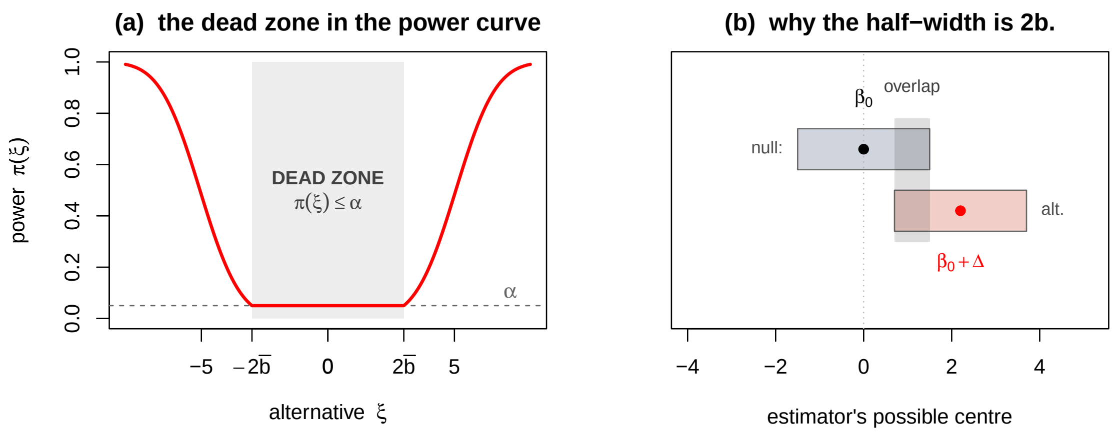

Figure 1: Left: the sharp worst-case power frontier, flat at α across the dead zone $|\xi| \leq 2\bar{b}$ and rising only outside. Right: the observational-equivalence picture, in which the bias can displace the observable centre of either world by up to $\bar{B}$, so the two worlds collide whenever their targets differ by less than $2\bar{B}$.

4 Main results

4.1 Three regimes: a race between $\bar{B}$ and se

Measure alternatives in standard-error units (the yardstick of the conventional test). The asymptotic outcome is decided entirely by the ratio $\bar{B}/se$:

regime	ratio $\bar{B}/se$	local power of the honest test
bound negligible	$\to 0$	conservatism asymptotically *free*
matched rate	$\to$ constant	dead zone of fixed width $2\bar{b}$; loss bounded, explicit
bound dominates	$\to \infty$	**zero** local power: no rival value is ever rejected

The third row is not exotic: it is the norm whenever a fixed, non-shrinking bound (a fixed M, a fixed $\bar{\gamma}$, a fixed number of weak factors) meets a $\sqrt{n}$-consistent estimator. There the honest interval's width is asymptotically all bias allowance, the dead zone engulfs every local alternative, and the test's local power function collapses to α everywhere.

4.2 No honest procedure escapes

One might hope a cleverer construction (a different estimator, an asymmetric interval, a data-dependent width) could dodge the zone. It cannot: a minimax theorem shows the dead zone is a property of *honesty itself*, not of any particular interval. The reason is the two-worlds collision above: within $2\bar{B}$, the rival world and the true world can generate identical data, and no procedure can tell apart identical distributions. The standard interval is, moreover, rate-optimal: honesty, not design, is what binds. At the matched rate the exact worst-case frontier is

$$\Phi\Big(\frac{(|\xi| - 2\bar{b})_+}{\sigma} - z_{1-\alpha}\Big),$$

flat at α on the zone and one-sided Gaussian beyond it: the textbook power envelope, translated outward by $2\bar{b}$.

4.3 IV and DiD need no asymptotics at all

In the two leading applied settings the collision is exact and elementary. In plausibly exogenous IV, the worlds $(\beta_0,\ \gamma = \bar{\gamma})$ and $(\beta_0 + \Delta,\ \gamma = \bar{\gamma} - \Delta\pi)$ generate *observationally identical* data for every $\Delta \leq 2\bar{\gamma}/\pi$, once the shift is absorbed into the structural error (a redefinition the model leaves free): not close but identical, at every sample size. In honest DiD, a linear violation of parallel trends of slope M paired with target β_0 produces the same event-study means as slope $M - \Delta$ paired with $\beta_0 + \Delta$, for every $\Delta \leq 2M$. Two consequences: the flat dead zone in these settings requires no regularity conditions and no limit theory; and *more data never helps*, because the bound is a belief, not a statistic. Concretely: with $\bar{\gamma} = 0.3$ and a first stage of $\pi = 1$, no honest test can be guaranteed to detect any violation of the null smaller than 0.6, forever.

4.4 Partial identification is the limiting case

A confidence interval's rescaled width and its local power are two views of one object: power against a rival value is the probability the limiting interval *excludes* it. Hence the organising claim: **local power degenerates exactly when the (suitably rescaled) interval width has positive limiting volume**, and partial identification, where the width never shrinks at all, is simply the extreme point $\delta = 0$ of the same phenomenon. Under partial identification, any rival value in the interior of the identified set is consistent with the same law as the truth, so honest power against it is at most α for *any* approach rate. Honest inference under an untestable bound is partial identification at a slower time-scale: the local identified set has diameter $2\bar{B}$, and the dead zone is its shadow in power.

5 Seeing it in simulations

Figure 2 shows the plausibly exogenous IV design at two extremes of the realised exclusion violation, with the bound fixed at $\bar{\gamma} = 0.3$ throughout ($\pi = 1$, so $\bar{b} = 0.3$ and the guarantee-free region spans 0.6). Each row sweeps rival values around the truth for $n = 100, 500, 2500$; blue dashes are the conventional IV test, red the honest (union-of-CIs) test, and the grey band is the realised dead window.

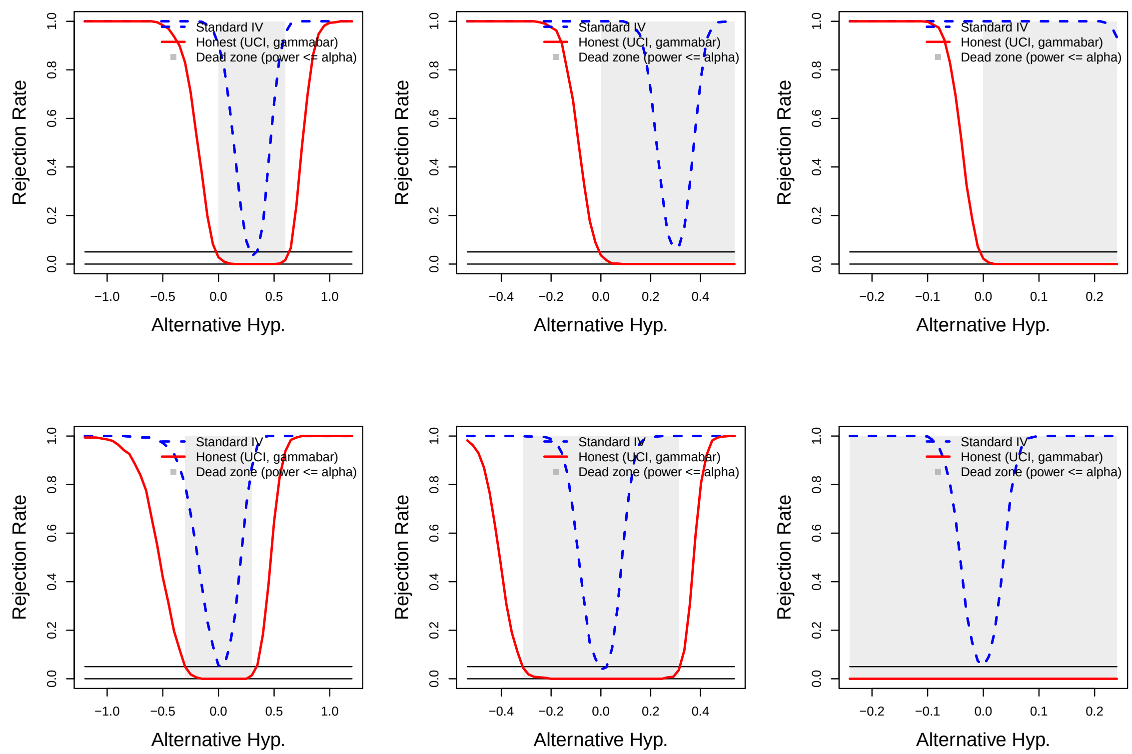

Figure 2: Honest power in plausibly exogenous IV, $\bar{\gamma} = 0.3$. Top: strongly endogenous instrument ($\gamma = 0.3$, bias at the bound). The window is one-sided, $[0, 2\bar{b}]$: its far edge sits at *twice* the bound from the truth, and the conventional test (blue) is badly oversized. The honest width is money well spent. Bottom: exogenous instrument ($\gamma = 0$, no bias). The window is the symmetric $\pm\bar{b}$; the conventional test is valid and sharp, and the honest test's entire sacrifice buys insurance against a bias that never occurred. Note the windows do not shrink with n in either row.

The two rows are the two halves of the two-number example of Section 2. The top row is the world where the budget was needed; the bottom row is the pure-waste world. The analyst cannot tell which row they are in (that is what "untestable" means) and so must pay the same width in both. The union of the two grey bands (and of all the bands in between) is the $\pm 2\bar{b}$ dead zone the paper says to disclose.

Figure 3 performs that union explicitly. Sweeping the realised violation over eleven admissible values traces out the family of sliding windows (thin grey), each of width $2\bar{b}$, centred wherever the bias happens to sit. Their lower envelope (black) is the worst case over the biases, which is what the dead zone's definition takes: it is flat on exactly $\pm 2\bar{b}$ and rises steeply outside. The red dashed curve is the *average* over a uniformly drawn bias: it ramps linearly and first reaches full power exactly at the zone's edge. So a randomly drawn bias locates $2\bar{b}$, but only the worst case is flat across it: the dead zone is a guarantee (an infimum), not an average.

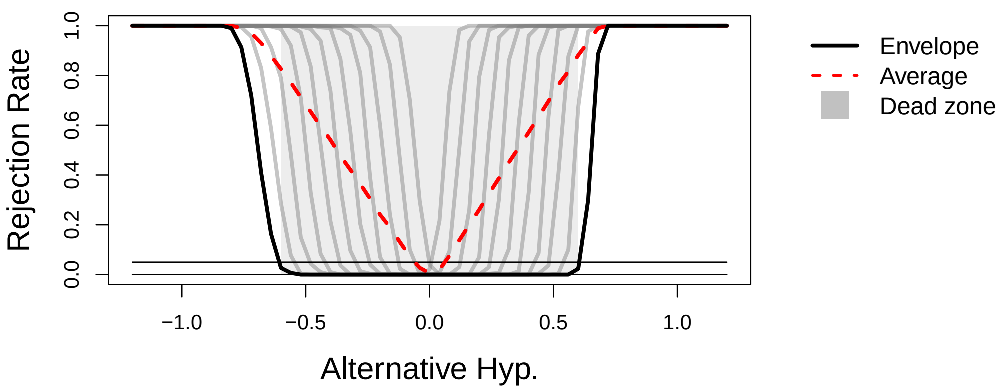

Figure 3: The union made visible: the same IV design with the realised violation swept over eleven admissible values ($n = 1000$). Thin grey: per-bias power curves (sliding windows). Black: their pointwise minimum, flat on exactly $\pm 2\bar{b} = \pm 0.6$. Red dashed: the average over a uniformly drawn bias, ramping to one at the zone's edge.

6 What to do in practice

The paper's recommendation is disclosure, not abstinence. Honest intervals buy a guarantee worth having; the proposal is to *price* it:

1. **Report the dead-zone half-width $2\bar{B}$ alongside any bias-aware interval**: in the estimand's units, next to the interval itself. It tells the reader which effect sizes the design can be guaranteed to detect and which it cannot, under the stated bound.

2. **Treat rejections inside the zone with care.** A rejection of a value within $2\bar{B}$ of the null is a rejection the procedure could not have been relied upon to deliver: under some admissible bias its probability was at most α.

3. **Sample-size translation.** At the parametric rate, matching the oracle's power against an effect Δ requires inflating the sample by the factor $\big(\Delta/(\Delta - 2\bar{B})\big)^2$: an effect three

bias-bounds away needs *nine times* the data, and an effect inside the zone cannot be reached at any sample size.

4. **The bound is the design choice.** Since the zone scales with $2\bar{B}$ and the data cannot shrink it, the substantive debate belongs where it is anyway conducted, over the plausibility of M, $\bar{\gamma}$, or the number of weak factors, now with its power consequences on the table.

The paper works two empirical applications: the Lee (2008) incumbency RD, where the matched rate makes honesty cheap but not free, and the Benzarti–Carloni (2019) VAT study under honest DiD, where a relative-magnitudes trend bound, which the data cannot shrink, puts economically relevant effect sizes inside the dead zone.

7 Where this sits in the literature

The *length* face of this phenomenon is classical. Bahadur and Savage (1956) showed that over a class rich enough to hide arbitrary contamination, honest inference about the mean is impossible outright; the analyst's a priori bound is exactly what truncates that impossibility to something finite. Donoho (1994) solved the resulting fixed-length interval problem and showed the honest length cannot fall below the bias bound; Low (1997) proved that even data-dependent lengths inherit the worst case of the class the analyst admits; Cai and Low (2004) built the complete adaptation theory for expected length. By the classical duality of Pratt (1961), length statements of this kind *imply* that power is lost somewhere, but not where, how much, or with what shape. This paper supplies the power face: the dead zone and its exact $2\bar{B}$ extent, the three regimes, the minimax theorem showing no honest procedure escapes, the sharp frontier and its attainment, and the unification with partial identification.

Paper. *Local Asymptotic Power of Honest Confidence Intervals*, available on request or via my website.