Logs with Zeros? Some Problems and Solutions

Jiafeng Chen & Jonathan Roth

The Quarterly Journal of Economics, Volume 139, Issue 2, May 2024, Pages 891–936.

Presented by Pauline Bucher - Thanks to Frank DiTraglia for digging for critics with me!.

Potential Outcomes Framework - Average Treatment Effect or ATE

When: Anytime you want to quantify the impact of an "intervention", e.g. job training programs, drug trials, smaller class sizes etc.

Potential Outcomes:

- $Y(1)$: Outcome if treated - e.g. wage after receiving the training.
- $Y(0)$: Outcome if not treated (control) - e.g. wage without receiving the training.

Average Treatment Effect (ATE):

$$ATE = \mathbb{E}[Y(1) - Y(0)]$$

The ATE represents the average difference in outcomes between the treated and untreated conditions across the population.

e.g. By how much the wages of an average individual in the treatment group increased compared to the control group.

Log Function and ATEs: Strictly Positive Outcomes

Interpretation:

- Interpretability: With $Y(0) > 0$ and $Y(1) > 0$, the ATE in logs approximates percentage changes:

$$\mathbb{E}[\log(Y(1)) - \log(Y(0))],$$

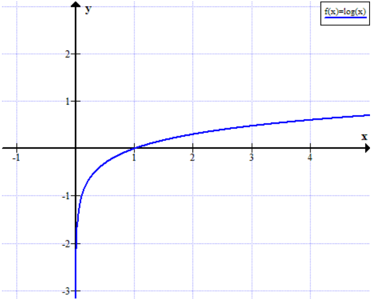
$$\log\left(\frac{Y(1)}{Y(0)}\right) = \log(Y(1)) - \log(Y(0)) \approx \frac{Y(1) - Y(0)}{Y(0)}$$

Intensive Margin: The ATE only measures the effect of the treatment on individuals who already have non-zero outcomes, e.g. the wage increase of those already working (non-zero).

Log Function and ATEs: Weakly Positive Outcomes

Problem:

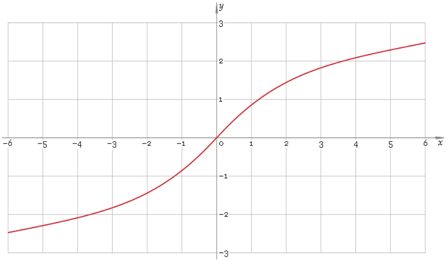
Many economic variables have $Y(1) > 0$ BUT $Y(0) = 0$, but $\log(0)$ is not defined.



(BAD) Common Solutions:

- Transformations like $\log(c + Y)$ to handle zeros.

- Alternative functions like the inverse hyperbolic sine function $\text{arcsinh}(Y)$.



Extensive Margin: The effect of the treatment on individuals who go from zero outcomes to positive ones, e.g. the wage going from unemployed to employed.

ATE - Extensive and Intensive Margins

Notation: $m(y)$:

- $m(y)$ behaves like $\log(y)$ for large y, e.g. the previous functions.
- Formally, we have:

$$\frac{m(y)}{\log(y)} \to 1 \quad \text{as} \quad y \to \infty$$

Potential Outcomes: Let $D \in \{0, 1\}$ and $Y \in [0, \infty)$.

$$Y = DY(1) + (1 - D)Y(0),$$

ATE Decomposition:

$$\text{ATE} = \underbrace{P(Y(1) > 0, Y(0) > 0) \cdot \mathbb{E}[m(Y(1)) - m(Y(0)) | Y(1) > 0, Y(0) > 0]}_{\text{Intensive Margin}}$$

$$+ \underbrace{P(Y(1) > 0, Y(0) = 0) \cdot \mathbb{E}[m(Y(1)) - m(0) | Y(1) > 0, Y(0) = 0]}_{\text{Extensive Margin}}$$

Scale Sensitivity of log ATEs ”with zeros”

- Proposition 1 (patience) establishes that with **weakly positive outcomes, the ATEs obtained via any m(Y) transformations will not be scale invariant and therefore cannot be interpreted as percentages**.

However...

- Between 2018-2022, 17 papers in the **American Economic Review (AER)** have used the **arcsinh(Y)** transformation to estimate treatment effects.
  - 10 explicitly interpret results as **percentage changes** or **elasticities**.
  - 6 do not directly interpret the units.

Scale Sensitivity of log ATEs ”with zeros”

Replication Study:

- Replicated results from 10/17 papers with publicly available data.
- Tested sensitivity to scaling by comparing **arcsinh(Y)** and **arcsinh(100 · Y)** (e.g., dollars vs. cents).

Key Finding:

- In **5 out of 10** articles, scaling the outcome by a factor of 100 changed treatment effects by more than **100%** of the original estimate.

Main Objectives of the paper

- Demonstrate the scale-sensitivity of log-like ATEs with zeros (Proposition 1)
- Show that the degree of sensitivity is positively correlated with the size of the extensive
- Establish a trilemma when trying to capture ATEs with weakly positive outcomes (Proposition 2)
- Suggest some question-dependent alternatives to log ATEs in those settings.

Literature Review

- Genesis: First use of log(1+Y) in 1937.
- Importantly, **Previous work has identified in simulations/empirical applications the sensitivity of some log transformations to the units of the outcome** (Aihounton et al., 2021) or (de Brauw et al., 2021).
- Mullahy and Norton (forthcoming) show that marginal effects in regressions with transformations like $\log(1+Y)$ or arcsinh(Y) are sensitive to outcome scaling, with marginal effects approaching that of a levels regression when units are small but that of a LPM when units are large.
  - This paper proves that scale dependence is necessary for any identified ATE that is well-defined with zero-valued outcomes + the dependence on units is arbitrarily bad for transformations that approximate log (Y) for large values of Y.

Literature Review

- A reoccurring theme in the paper and its Appendix is a refusal of the alternative to detangle the extensive/intensive margins with zero-valued outcomes, namely, parametric models like **Tobin (1958) or Heckman (1979)**.
- These methods impose parametric structure on the joint distribution of the potential outcomes, which the authors denote as often "hard to justify". This paper shows what can be done with zero-valued outcomes without any restriction on the the joint distribution

Carranza et al. (2022): Set-Up

- Randomized Controlled Trial (RCT) in South Africa:
  - Individuals in the treatment group received certified test results that they **can** show to prospective employers to vouch for their skills.
- **Outcome of Interest:** Weekly Hours Worked

$$\text{arcsinh}(Y_i) = \beta_0 + D_i\beta_1 + X_i'\gamma + u_i$$

- **Result:** The estimated **ATE, $\hat{\beta}_1$, is 0.201, which is interpreted as a 20% increase in weekly hours worked for those with certification**.

Proposition 1: Sensitivity of Log ATEs with Zeros

Set-up:

- We consider average treatment effects (ATEs) of the form:

$$\theta = \mathbb{E}_P\left[m(Y(1)) - m(Y(0))\right]$$

- θ represents the ATE among the population indexed by P. If P refers to a subpopulation (e.g., compliers for an instrument), then θ is the local average treatment effect (LATE).

- We are interested in how $\theta(a)$ changes if we scale the outcome Y by a factor a:

$$\theta(a) = \mathbb{E}_P\left[m(aY(1)) - m(aY(0))\right]$$

The proof (available in the Online Appendix) relies on first showing the continuity of the function $\theta(a)$ and then using the Intermediate Value Theorem.

Assumptions:

- The function m is continuous and increasing, $m : [0, \infty) \to \mathbb{R}$.
- $m(y)/\log(y) \to 1$ as $y \to \infty$, meaning the function m behaves like $\log(y)$ for large values of y.
- The treatment affects the extensive margin: $P(Y(1) = 0) \neq P(Y(0) = 0)$.
- Finite expectations [1]: $\mathrm{E}_{PY(d)}\left[|\log(Y(d))| \mid Y(d) > 0\right] < \infty \quad$ for $d = 0, 1$.

- **Then, for any $\theta^* \in (0, \infty)$, there exists an $a > 0$ such that $|\theta(a)| = \theta^*$.**
- **As $a \to 0$, $\theta(a) \to 0$, and as $a \to \infty$, $\theta(a) \to \infty$.**

Fret none: When outcomes are both strictly positive (no extensive margin), if $m(y) = \log(y)$, then $\theta(a)$ is indeed scale-invariant, meaning the ATE in logs does not depend on the value of a.

[1] This assumption simply ensures that $E_{Y(d)}\left[|m(aY(d))| \mid Y > 0\right]$ exists for all values of $a > 0$.

Scale Invariance

A function is said to be **homogeneous of degree 0** if scaling all inputs by a constant factor doesn't change the function's value.

$$f(c \cdot x) = f(x) \quad \text{for any constant } c$$

In this paper, the sensitivity to the scale of the outcome means the

ATEs are **not HOD 0**. Rescaling (e.g., dollars to cents) can change the estimated treatment effect.

Remark 2: Finite Changes in Scaling

- **Proposition 1** shows that any magnitude of $|\theta(a)|$ can be achieved via the appropriate choice of a.
- **Question:** How much does $\theta(a)$ change for finite changes in a?
- **Proposition 4 (Online Appendix):** For large scaling factors a, the change in ATE is approximately:

$$E_P[m(aY(1))-m(aY(0))] = (P(Y(1)>0)-P(Y(0)>0))\cdot\log(a)+o(\log(a)$$

- The ATE for $m(Y)$ is more sensitive to finite changes in scaling a when the extensive-margin effect is large.[2]
- **Implication:** Sensitivity analysis under finite changes in c in $\log(c+Y)$ captures the size of the extensive margin.
- Estimate extensive-margin effect by using the same procedure as in the original paper but with the outcome $\mathbf{1}[Y>0]$.

[2] P(one in the T group has a positive outcome) - P(one in the control group has a positive outcome) = increase in prob of a positive outcome due to treatment.

Summary of the issue

- In 5/10 articles, scaling the outcome by a factor of 100 changed treatment effects by more than 100% of the original estimate.
- Sensitivity was low (less than 10%) in papers with near-zero extensive margin effects.

TABLE I

CHANGE IN ESTIMATED TREATMENT EFFECTS FROM RESCALING THE OUTCOME BY A FACTOR OF 100 IN ARTICLES PUBLISHED IN THE *AER* USING arcsinh(Y)

	Treatment effect using:			Change from rescaling units:	
Paper	arcsinh(Y)	arcsinh($100 \cdot Y$)	Ext. margin	Raw	%
Azoulay et al. (2019)	0.003	0.017	0.003	0.014	464
Fetzer et al. (2021)	−0.177	−0.451	−0.059	−0.273	154
Johnson (2020)	−0.179	−0.448	−0.057	−0.269	150
Carranza et al. (2022)	0.201	0.453	0.055	0.252	125
Cao and Chen (2022)	0.038	0.082	0.010	0.044	117
Rogall (2021)	1.248	2.150	0.195	0.902	72
Moretti (2021)	0.054	0.068	0.000	0.013	24
Berkouwer and Dean (2022)	−0.498	−0.478	0.010	0.020	−4
Arora et al. (2021)	0.113	0.110	−0.001	−0.003	−3
Hjort and Poulsen (2019)	0.354	0.354	0.000	0.000	0

Role of the Size of the Extensive Margin

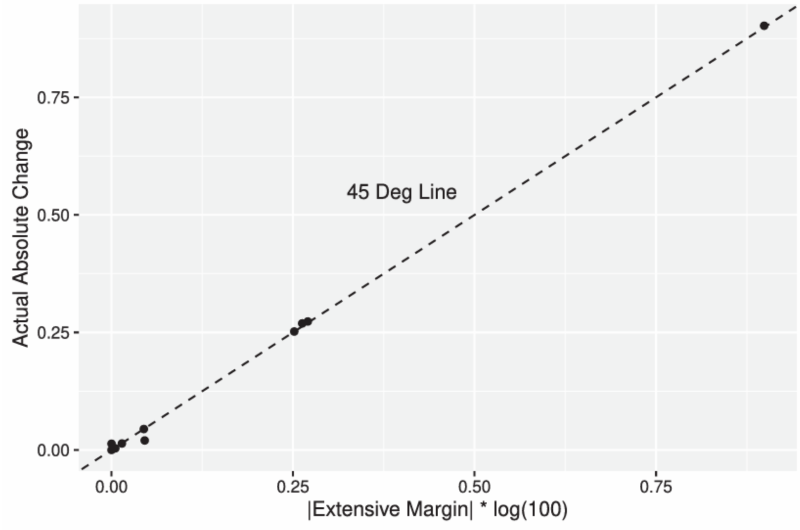


FIGURE I

Other remarks

- **The scale-sensitivity result extends to continuous treatment** [3] whenever the treatment effect contrast averages $m(aY(d))$ across possible values of d, it will be sensitive to scaling when there is an extensive margin effect.
- **Scale Sensitivity of OLS estimands even when they do not have a causal interpretation:** The results are about estimands thus any consistent estimator of the ATE for m(Y) will be sensitive to scaling (at least asymptotically) when there is an extensive margin.
- **The t-statistic for $\theta(a)$, i.e., estimated via $\hat{\beta}_1$ in regressions using log-like transformation of the outcome, converges to that for the extensive-margin effect estimate** when replacing the LHS with the indicator function $1[Y > 0.]$

[3] Consider a study examining the impact of hours of training on job performance.

Carranza et al. (2022): Problem

- Outcome: Weekly Hours Worked, **estimated ATE, $\hat{\beta}_1$, is 0.201**

$$\text{arcsinh}(Y_i) = \beta_0 + D_i\beta_1 + X_i'\gamma + u_i$$

- i) Yearly hours worked = weekly hours x 52
- ii) Full Time Equivalents worked = weekly hours / 40

TABLE III

ESTIMATES USING arcsinh(Y) WITH DIFFERENT UNITS OF Y IN CARRANZA ET AL. (2022)

	arcsinh(weekly hrs) (1)	arcsinh(yearly hrs) (2)	arcsinh(FTEs) (3)
Treatment	0.201 (0.052)	0.417 (0.096)	0.031 (0.012)
Units of outcome:	Weekly hrs	Yearly hrs	FTEs

Definitions

Point Identification from the Marginals at P:

- Let P denote the joint distribution of $(Y(1), Y(0))$ and let $P_{Y(d)}$ denote the marginal distribution of $Y(d)$ for $d = 0, 1$.
- We say that θ_g is **point identified** from the marginals at P if for every joint distribution Q with the same marginals as P, i.e., such that:

$$Q_{Y(d)} = P_{Y(d)} \quad \text{for } d = 0, 1,$$

  we have:

$$\mathbb{E}_P[g(Y(1), Y(0))] = \mathbb{E}_Q[g(Y(1), Y(0))].$$

- For a class of distributions $\mathcal{P}$, we say that θ_g is point identified over $\mathcal{P}$ if for every $P \in \mathcal{P}$, θ_g is point identified from the marginals at P.

Conclusion: Point identification ensures that the treatment effect estimate θ_g is consistent across distributions with the same marginals, even if the joint distribution varies.

Proposition 2: A Trilemma

Proposition 2 states that the following three properties cannot hold simultaneously when there are zero-valued outcomes:

1. $\theta_g = \mathbb{E}_P[g(Y(1), Y(0))]$ **for a nonconstant function** $g : [0, \infty)^2 \to \mathbb{R}$**, which is weakly increasing in its first argument.**
2. **The function g is scale invariant.**
3. θ_g **is point identified over** $\mathcal{P}_+$ **(the set of distributions on** $[0, \infty)$**).**

Implications:

One cannot "fix" the issues with ATEs for log-like transformations by taking alternative transformations of the outcome (e.g., $\sqrt{Y}$).

If $g(y_0, y_1) = m(y_1) - m(y_0)$ and $m(y)$ is an increasing function, the ATE for $m(Y)$ cannot be scale invariant while also being point identified.

Implications for Settings without an Extensive Margin

When outcomes are strictly positive...

- The only parameter satisfying all 3 properties is the ATE in logs.
- **Any other transformation of the outcome other than log(Y) will depend on the units of the outcome for at least some data-generating process (DGP).**
- However, if there is no extensive margin, the ATE for a log-like transformation will be approximately insensitive to scaling once the values of Y are large.

Second Best solutions

TABLE II
SUMMARY OF ALTERNATIVE TARGET PARAMETERS

Description	Parameter	Main property sacrificed?	Pros/cons
Normalized ATE	$\frac{E[Y(1)-Y(0)]}{E[Y(0)]}$	$E[g(Y(1), Y(0))]$	Pro: Percent interpretation Con: Does not capture decreasing returns
Normalized outcome	$E\left[\frac{Y(1)}{X} - \frac{Y(0)}{X}\right]$	$E[g(Y(1), Y(0))]$	Pro: Per-unit-X interpretation Con: Need to find sensible X
Explicit trade-off of intensive/extensive margins	ATE for $m(y) = \begin{cases} \log(y) & y > 0 \\ -x & y = 0 \end{cases}$	Scale invariance	Pro: Explicit trade-off of two margins Con: Need to choose x; monotone only if support excludes $(0, e^{-x})$
Intensive-margin effect	$E\left[\log\left(\frac{Y(1)}{Y(0)}\right) \mid Y(1) > 0, Y(0) > 0\right]$	Point identification	Pro: ATE in logs for the intensive margin Con: Partial identification

Normalized ATE in levels

When? Whenever the goal is easily interpretable units (e.g. %).

- One approach is to target the ATE in levels expressed as a % of the control mean :

$$\theta_{\text{ATE\%}} = \frac{E[Y(1) - Y(0)]}{E[Y(0)]}$$

- For example, if a researcher is studying a program D designed to reduce healthcare spending Y, then $\theta_{\text{ATE\%}}$ represents the percentage reduction in costs resulting from the program.
- $\theta_{\text{ATE\%}}$ is point identified and scale-invariant but it cannot be written as $\theta_g = \mathbb{E}_P[g(Y(1), Y(0))]$.
- Importantly, $\theta_{\text{ATE\%}}$ captures the percentage change in the average outcome between treatment and control, but it is not an average of individual-level percentage changes[4]

[4] It gives us an aggregate view.

Normalized ATE in Levels

- $\theta_{ATE\%}$ can be consistently estimated using **Poisson regression** under appropriate assumptions.
- The Poisson quasi-maximum likelihood (QMLE) approach[5]: With a randomly assigned D, estimation of $Y = \exp(\alpha + \beta D)U$ estimates the population coefficient β, where:

$$e^{\beta} - 1 = \frac{E[Y(1)]}{E[Y(0)]} - 1 = \theta_{ATE\%}$$

  - Does not distinguish between E/I margins (OK for count data e.g. nb of publications, less for things like job training programs).
  - Just a rescaling of the ATE in levels, may be dominated by individuals in the tail of the distribution.

[5] Wooldridge (2010); Santos Silva and Tenreyro (2006)

Carranza et al. (2022): Percentage Changes in the Average

- **Average weekly hours worked**: 9.84 for the Treated Group and 8.85 for the Control one.
- Estimate of $\theta_{\text{ATE\%}} = \frac{E[Y(1)-Y(0)]}{E[Y(0)]}$, 9.84/8.85= 1.11.
- Estimating using a Poisson specification (allows covariates):

$$Y_i = \exp(\beta_0 + \beta_1 D_i + X_i'\gamma)U_i$$

$$\hat{\theta}_{\text{ATE\%}} = \exp(\hat{\beta}_1) - 1 = 0.11$$

TABLE IV

POISSON REGRESSION AND IMPLIED PROPORTIONAL EFFECTS IN CARRANZA ET AL. (2022)

	(1)	(2)
β_0	2.180	1.291
	(0.058)	(0.311)
β_1	0.106	0.140
	(0.072)	(0.060)
Implied prop. effect	0.112	0.150
	(0.081)	(0.069)
Covariates	No	Yes

Normalized Outcome

When? Whenever the goal is easily interpretable units (e.g. percentages).

- One can also consider an outcome of the form:

$$\tilde{Y} = \frac{Y}{X}$$

  where Y is the original outcome, and X is some predetermined characteristic (e.g., population for employment-to-population ratio).
- The ATE for $\tilde{Y}$ is scale-invariant, point-identified but cannot be written as

$$\tilde{\theta} = E_P[g(Y(1), Y(0))]$$

- Requires having such a variable X. Typically can use pretreatment observations of the outcome (assuming these are positive), or the predicted control outcome given some observable characteristics.

Normalized Outcome

- A second example is to use $\tilde{Y} = F_{Y^*}(Y)$
  - The transformed outcome $\tilde{Y}$ corresponds to the rank (percentile) of an individual in the reference distribution.
  - The ATE for $\tilde{Y}$ is unit invariant as long as Y and Y^* are measured in the same units.
  - This approach has been found to yield more stable estimates compared to log transformations, especially in the context of intergenerational mobility (i.e. where Ytilda corresponds to a child's rank in the national income distribution).

- Another option: transform outcomes of the form $1[Y > y]$ for different values of y to report treatment effects on specific thresholds (e.g., probability of earning at least $50,000).
  - Combines the effect of the treatment along both the intensive and extensive margins- for example, a worker who has $Y(1) > \$50,000 > Y(0)$ could either not work under control $Y(0) = 0$ or work under control but have earnings below $50,000.

Explicit Trade-Off of Intensive/Extensive Margins

When? Whenever one wants to captures decreasing returns.

- For example, when Y is strictly positively valued, the **ATE in logs** corresponds to the change in utility from implementing the treatment for a **utilitarian social planner** with log utility:

$$U = E[\log(Y)]$$

Challenges:

- **Not well-defined** when there is an extensive margin. Scaling of the outcome implicitly determines the weights placed on these margins.

$\rightarrow$ By proposition 2, we won't have scale invariance. A transparent approach is to explicitly take a stand on **how much one values the change from 0-1 relative to a percentage change** in earnings for nonzero outcomes. DON'T weigh the margins via the scaling of Y.

Explicit Trade-Off Intensive/Extensive Margins

- If one values the **extensive margin effect** of moving from 0 to 1 the same as a 100x% increase in earnings, consider setting[6] :

$$m(y) = \log(y) \quad \text{for } y > 0, \quad m(0) = -x$$

- The ATE for this transformation can be interpreted as an **approximate percentage (log-point) effect**:

Increase from 0 to 1 valued at $100x$ log points.

- Still debate over appropriate x or utility functions.

Dependence on scaling may be a plus here...The appropriate choice of x also depends on the units of the outcome—for example, saying a change from 0 to 1 is worth 100 x means something very different if 1 corresponds with $1 versus $1 million.

[6] Footnote 28: Note that this transformation will generally only be sensible if the support of Y excludes (0, e^{-x}), because otherwise the function m (y) is not monotone in y over the support of Y . It is common, however, to have a lower bound on nonzero values of the outcome; for example, a firm cannot have between 0 and 1 employees.

Intensive Margin Effect: Problem

When? Whenever the goal is easily to separate both margins.

- **Card et al, 2010**: Whether a program raises participants' earnings by helping them find a job—which would be expected only to have an extensive- margin effect—or by increasing human capital, which would be expected to also affect the intensive margin.
- The $\theta_{\text{Intensive}}$ parameter captures the **ATE** in logs for those individuals who have a positive outcome, regardless of treatment status.

$$\theta_{\text{Intensive}} = E[\log(Y(1)) - \log(Y(0)) \mid Y(1) > 0, Y(0) > 0]$$

- This parameter is scale-invariant and can be written in as a ratio of 2 functions following $\theta_g = \mathbb{E}_P[g(Y(1), Y(0))]$.
- Hence by Proposition 2 it is **not Point Identified**.

Intensive Margin Effect: Solution(s)

- **Lee (2009)** popularized a method for obtaining **bounds** on $\theta_{\text{intensive}}$ under the **monotonicity assumption** [7]:
  - The assumption implies that everyone with positive earnings without receiving the treatment would also have positive earnings when receiving it.
- These bounds can be reported alongside the change in probability of having a nonzero outcome (extensive margin effect):

$$P(Y(1) > 0) - P(Y(0) > 0).$$

- The Lee (2009) bounds will tend to be tight when the extensive-margin effect is close to zero.

With Further Assumptions...:

- One can further tighten the bounds (or restore point identification) by imposing assumptions on the **joint distribution** of the potential outcomes.

[7] Footnote 30: See also Zhang and Rubin (2003) for related results, including bounds without the monotonicity assumption.

Carranza et al. (2022): Intensive Margin

- Estimated extensive-margin treatment effect of 0.055: **increases the fraction of people with positive hours worked by 5.5 percentage points**
- We may be interested in whether the overall 11% increase in hours worked is driven entirely by the extensive margin or not. Does the treatment increase hours only by bringing people into the labor force, or does it also allow those who would have worked anyway to find jobs with more hours (e.g., full-time instead of part-time)?

Carranza et al. (2022): Intensive Margin

- Using ”Lee Bounds” (2009) allows the computation of the treatment effect for ”always-takers”—individuals who would have positive hours worked regardless of treatment ($Y(1) > 0, Y(0) > 0$).
- **ACHTUNG Monotonicity Assumption:** Assumes that anyone who would work positive hours without treatment would also work positive hours when treated, i.e., $P(Y(1) = 0, Y(0) > 0) = 0$. Reasonable if workers only use certification when it improves job prospects.
- May want to assume more to increase precision, e.g. We might reasonably expect that the compliers are negatively selected relative to the always-takers and thus would work fewer hours when receiving treatment.

$$E[Y(1)|\text{Complier}] = (1 - c)E[Y(1)|\text{Always-taker}]$$

where c represents the proportional reduction in average hours for compliers relative to always-takers.

Carranza et al. (2022): Intensive Margin

- The bounds are large and include 0s for both (1) and (2).
- c = 0 (3): -1.02 weekly hours (negative intensive margin).
- c = 0.25 (4): Compliers work 25% fewer hours than always-takers $\rightarrow$: -0.07 weekly hours (no significant intensive margin).
- c = 0.5 (5): +0.95 weekly hours (positive intensive margin).

TABLE V

BOUNDS AND POINT ESTIMATES FOR THE INTENSIVE-MARGIN TREATMENT EFFECT IN CARRANZA ET AL. (2022)

	(1)	(2)	(3)	(4)	(5)
Lower bound	−0.195	−6.665			
	(0.064)	(1.366)			
Upper bound	0.283	2.771			
	(0.114)	(2.067)			
Point estimate			−1.025	−0.069	0.954
			(1.182)	(1.349)	(1.588)
Units	Log(hours)	Hours	Hours	Hours	Hours
c			0	0.25	0.5

Other Empirical Applications

- Besides Carranza et al. (2022), the authors follow the same replication/change of scale for the outcome to test the robustness of the estimated ATEs/try alternative parameters/solutions in two other contexts, namely **a Diff-in-Diff setting (Sequeira, 2016) and IV (Berkouwer et al, 2022)**.
- The results are very similar to that of Carranza et al. (2022) and in line with the proposition 1 and 2: Yes, with zero-valued outcomes, change of scales with log-like transformations are not scale invariant. Poisson regression yield great estimates of Normalised ATEs.
- The solution where one chooses how to value the extensive margin is in the DiD application. As expected, the estimated treatment effect grows in magnitude as we place more value on the extensive margin.

Discussion points

- Very practical, constant discussion of empirical examples. **BUT** This paper is not really that novel - its popularity feels largely due to its slickness. It is an accessible read that clearly addresses issues plaguing even excellent (AER standard) publications - you've now been warned and can do better!

- Great at showing issues (spared you the details of the other 2 applications) and proposes some fixes but often these rely on a few further assumptions too (e.g. for Lee Bounds, one needs monotonicity) or they are already largely implemented anyways (e.g. Normalized ATE). They feel very context-dependent. Yet the paper does not discuss/try writing structural models[8].

- All in all, **"Logs with Zeros?" Some *already discussed, but better here, problems* Problems, and Some *but don't expect too much/we don't like structural models* Solutions.**

[8] Andrew Gelman has a couple of blog posts, published before this paper, on the issue. There are great discussions in the comments of various fixes https://statmodeling.stat.columbia.edu/2024/08/31/loga-x-not-log1-x/