

Corrigendum: Logs with zeros? Some problems and solutions

Deniz Okat
University of Technology Sydney

Jiafeng Chen
Stanford University

Jonathan Roth
Brown University

October 12, 2026

Abstract

This corrigendum corrects an error in the statement of Proposition 2 of [Chen and Roth \(2024\)](#). The published version of Proposition 2 states that when the outcome $Y \geq 0$ can take the value zero, there is no treatment effect parameter that (a) is an average of individual level effects of the form $E[g(Y(1), Y(0))]$, where $g : [0, \infty)^2 \rightarrow \mathbb{R}$ is a non-constant function weakly increasing in $Y(1)$, (b) is scale-invariant, and (c) is point-identified from the marginal distributions of the potential outcomes. For the proposition to be correct, condition (a) needs to be strengthened to assert that the function g is non-constant *on the strictly positive real domain*, $(0, \infty)^2$, so that the function g puts non-zero weight on intensive margin changes. Without this condition, the percentage point change in the extensive margin, $\theta^{EM} = E[1\{Y(1) > 0\} - 1\{Y(0) > 0\}]$, satisfies all three conditions of the proposition. The practical recommendations of the paper are unchanged.

This corrigendum corrects an error in the statement of Proposition 2 of [Chen and Roth \(2024\)](#). The published version of the proposition reads:

Proposition 2 (A trilemma). *The following three properties cannot hold simultaneously:*

- (a) $\theta_g = E_P[g(Y(1), Y(0))]$ for a non-constant function $g : [0, \infty)^2 \rightarrow \mathbb{R}$ that is weakly increasing in its first argument.
- (b) The function g is scale-invariant.
- (c) θ_g is point-identified over $\mathcal{P}_+$.

However, Proposition 2 is not correct as stated. The treatment effect on the extensive margin, $\theta^{EM} := E[1\{Y(1) > 0\} - 1\{Y(0) > 0\}]$, clearly satisfies all three properties. The statement of Proposition 2 is correct if one strengthens condition (a) to require that the function g is non-constant on $(0, \infty)^2$. A corrected version of the Proposition is as follows:

Corrected Proposition 2 (A trilemma). *The following three properties cannot hold simultaneously:*

- (a) $\theta_g = E_P[g(Y(1), Y(0))]$ for a ~~non-constant~~ function $g : [0, \infty)^2 \rightarrow \mathbb{R}$ that is ~~non-constant on~~ $(0, \infty)^2$ and weakly increasing in its first argument.

(b) The function g is scale-invariant.

(c) θ_g is point-identified over $\mathcal{P}_+$.

We briefly explain the error in the proof of Proposition 2. Proposition 3 of the paper establishes that if $g : (0, \infty)^2 \rightarrow \mathbb{R}$ is scale-invariant and weakly increasing in the first argument, and $E[g(Y(1), Y(0))]$ is point-identified for all marginal distributions over the strictly-positive reals, then $g(Y(1), Y(0)) = c \log(Y(1)/Y(0)) + d$ for constants $c \geq 0, d$. The published proof of Proposition 2 asserts that if $g : [0, \infty)^2 \rightarrow \mathbb{R}$ satisfies conditions (a)-(c) of Proposition 2, then g must equal $c \log(Y(1)/Y(0)) + d$ for $c > 0$ on $(0, \infty)^2$, and then derives a contradiction. This assertion is incorrect, because a function that is non-constant over $[0, \infty)^2$ can be constant over $(0, \infty)^2$, and hence it is possible that $c = 0$. The published proof is therefore correct if one adds the assumption that the function g is non-constant on $(0, \infty)^2$, in which case by Proposition 3, $g(Y(1), Y(0)) = c \log(Y(1)/Y(0)) + d$ on $(0, \infty)^2$ for $c > 0$, and the contradiction derived in the published proof applies.

In Lemma 1 below, we show that the counterexamples to the originally-stated Proposition 2 are limited to parameters of the form $\theta_g = \alpha + \beta E[1\{Y(1) > 0\}] + \gamma E[1\{Y(0) > 0\}]$, with θ^{EM} corresponding to the special case $\alpha = 0, \beta = 1, \gamma = -1$.

Takeaways. Chen and Roth regret the error in the original article and are grateful to Okat for bringing it to their attention. Nevertheless, we think the practical takeaways from Chen and Roth (2024) remain unchanged by correcting the statement of Proposition 2. Proposition 2 of Chen and Roth (2024) aimed to establish that it is impossible to point-identify an average of individual-level effects with a *percentage* effect interpretation when the outcome can take on zeros. Although $\theta^{EM} = E[1\{Y(1) > 0\} - 1\{Y(0) > 0\}]$ is scale invariant, it should be interpreted as the treatment's impact on the probability that the outcome is non-zero, rather than the percentage effect of the treatment on the outcome. In fact, θ^{EM} is equal to zero even if the treatment, say, doubles the outcome for all units, since it places no weight on the intensive margin. The revised statement of Proposition 2 simply removes from consideration parameters like θ^{EM} that place no weight on intensive margin changes, and thus cannot reasonably be interpreted as measuring percentage changes in the outcome. The practical implications are therefore the same: in settings with zero-valued outcomes, researchers must still give up on one of the arms of the (corrected) trilemma. We think the approaches outlined in Section 4 of Chen and Roth (2024), which consider which arm a researcher should relax depending on their goals, remain useful. Finally, we note that one of the recommendations in Chen and Roth (2024) is indeed to report θ^{EM} alongside estimates of (bounds on) the intensive margin treatment effect. For example, in an application to Carranza et al. (2022), Chen and Roth (2024) report estimates of θ^{EM} along with estimates of bounds on the intensive margin treatment effect for individuals with positive values of Y under both treatments, $E[\log(Y(1)/Y(0)) \mid Y(1) > 0, Y(0) > 0]$.

Formal characterization of counterexamples. The following lemma characterizes the set of counterexamples to Proposition 2 as originally stated.

Lemma 1. *A parameter θ_g satisfies conditions (a)-(c) of Proposition 2 as originally stated if and only if it takes the form*

$$\theta_g = \alpha + \beta E[1\{Y(1) > 0\}] + \gamma E[1\{Y(0) > 0\}], \quad (1)$$

where $\beta \geq 0$ and at least one of β, γ is non-zero.

Proof. It is immediate that all parameters of this form satisfy conditions (a)-(c) as originally stated. We now show that all parameters satisfying the original (a)-(c) take this form. The Corrected Proposition 2 establishes that no parameter can satisfy the originally-stated (a)-(c) if the function $g : [0, \infty)^2 \rightarrow \mathbb{R}$ is non-constant on $(0, \infty)^2$. Thus, suppose that $g(y_1, y_0) = k$ for all $(y_1, y_0) \in (0, \infty)^2$.

We claim that if θ_g is identified over $\mathcal{P}_+$, then $y_1 \mapsto g(y_1, 0)$ is constant on $(0, \infty)$. To show this, consider distinct $y_1, y'_1 > 0$. Let $P_1 = \frac{1}{2}\delta_{(y_1, 0)} + \frac{1}{2}\delta_{(y'_1, y_1)}$ be the two-point distribution that randomizes $(Y(1), Y(0))$ with equal probability between $(y_1, 0), (y'_1, y_1)$ and analogously let $P_2 = \frac{1}{2}\delta_{(y'_1, 0)} + \frac{1}{2}\delta_{(y_1, y_1)}$ randomize with equal probability between $(y'_1, 0), (y_1, y_1)$. It is clear that P_1 and P_2 have the same marginals for $Y(1)$ and $Y(0)$. Identification of θ_g over $\mathcal{P}_+$ then implies that

$$E_{P_1}[g(Y(1), Y(0))] = \frac{1}{2}(k + g(y_1, 0)) = E_{P_2}[g(Y(1), Y(0))] = \frac{1}{2}(k + g(y'_1, 0))$$

and hence $g(y_1, 0) = g(y'_1, 0)$. Analogously, $y_0 \mapsto g(0, y_0)$ is constant on $(0, \infty)$.

Let $\alpha := g(0, 0)$, let β be the unique value of $g(y_1, 0) - g(0, 0)$ for $y_1 > 0$, let γ be the unique value of $g(0, y_0) - g(0, 0)$ for $y_0 > 0$, and let $\Delta := k - \alpha - \beta - \gamma$. It follows that

$$g(y_1, y_0) = \alpha + \beta 1\{y_1 > 0\} + \gamma 1\{y_0 > 0\} + \Delta 1\{y_1 > 0, y_0 > 0\}.$$

Next, we claim that if θ_g is identified over $\mathcal{P}_+$, then $\Delta = 0$. Indeed, consider $P_3 = \frac{1}{2}\delta_{(0, 0)} + \frac{1}{2}\delta_{(1, 1)}$ and $P_4 = \frac{1}{2}\delta_{(0, 1)} + \frac{1}{2}\delta_{(1, 0)}$. Since P_3, P_4 have the same marginals, identification over $\mathcal{P}_+$ implies

$$E_{P_3}[g(Y(1), Y(0))] = \alpha + \frac{1}{2}(\beta + \gamma + \Delta) = E_{P_4}[g(Y(1), Y(0))] = \alpha + \frac{1}{2}(\beta + \gamma),$$

which implies $\Delta = 0$. Since $\Delta = 0$, if g is weakly increasing in its first argument and non-constant, then $\beta \geq 0$ and at least one of β, γ is non-zero. Thus, $g(y_1, y_0) = \alpha + \beta 1\{y_1 > 0\} + \gamma 1\{y_0 > 0\}$ for $\beta \geq 0$ and at least one of β, γ is non-zero. This completes the proof. $\square$

References

Carranza, Eliana, Robert Garlick, Kate Orkin, and Neil Rankin. 2022. “Job Search and Hiring with Limited Information about Workseekers’ Skills.” *American Economic Review*, 112(11): 3547–83.

Chen, Jiafeng, and Jonathan Roth. 2024. “Logs with Zeros? Some Problems and Solutions.”
The Quarterly Journal of Economics, 139(2): 891–936.