What can a platform-wide AI rollout establish about sales?

Lluis Eriksson

20 November 2026

Abstract

A platform can compare how sales change across product categories even when every category has access to AI. Can those comparisons establish that total sales increased? Using Barwick et al.'s exposure design, we give the exact aggregate break-even condition under known relative effects on $\log(1 + \text{sales})$. The same observed panel and parallel untreated trends can imply a gain, zero, or a loss. A separate result translates a change in the historical-control calibration into a bound on the matched regression coefficient. Applied to the sales point in Figure III, a recalibration of roughly 42% or more would be needed to erase that estimate under nonnegative control weights of total mass at most one.

A gradient and a platform total

In the continuous-treatment literature [2], exposure contrasts and level effects are distinct. Barwick et al. [1] study a Chinese e-commerce platform that piloted AI in March 2024 and introduced its content and business-operation functions in April and July. Their Eq. (7) compares category outcomes by pre-pilot refund rate, including category and calendar-month fixed effects. The authors explicitly limit the coefficient to differential changes. Their 51.6% estimate for adopting firms answers a separate question. Proposition 3 of their structural model gives conditions for sales to rise, including a demand elasticity above one. We examine what the exposure comparison alone implies.

Fix a month and a finite set of disjoint sales categories covering the target population. Write y_i for observed sales and y_i^0 for sales absent the rollout. Grant more than a regression slope establishes: suppose the relative effects d_i are known exactly and

$$\log(1 + y_i) - \log(1 + y_i^0) = a + d_i.$$

For instance, $d_i = \beta(r_i - r_*)$ for refund rate r_i and reference r_*. Put $B = \sum_i (1+y_i)$, $w_i = (1+y_i)/B$ and $H = \sum_i w_i e^{-d_i}$. Solving for each counterfactual gives

$$\Delta = \sum_i (y_i - y_i^0) = B(1 - He^{-a}).$$

Thus expansion requires and is implied by $a > \log H$. The weights include the unit offset in the original currency units. Changing the reference shifts the common effect and the threshold equally. The formula requires effects on the category outcomes themselves; an average fitted log contrast alone cannot supply them.

Sharp bounds and an observational equivalence

Nonnegative counterfactual sales impose $a \le U = \min_i\{\log(1+y_i) - d_i\}$. Conversely every $a \le U$ constructs nonnegative counterfactuals by the displayed log equation. For real-valued sales the sharp range is

$$(-\infty,\ B(1 - He^{-U})].$$

An admissible external lower bound L on a changes the lower endpoint to $B(1 - He^{-L})$, attained at L. A zero-sales category at maximal d_i implies $U \le -\max d_i \le \log H$, which rules out a positive aggregate effect within this exact-shift model.

Consider two categories with pre-period sales $(119, 59)$ and observed post-period sales $(119, 119)$. Set $e^{d_1} = 1$, $e^{d_2} = 2$. Then $w_1 = w_2 = 1/2$ and $H = 3/4$. Each row provides a full set of untreated post-period outcomes; the untreated gross changes in $1+y$ are equal across categories within each row.

e^a	Untreated post sales	Common untreated growth factor	Effect Δ
1/2	$(239, 119)$	2	-120
3/4	$(159, 79)$	4/3	0
1	$(119, 59)$	1	$+60$

All observed sales and relative log effects are identical across rows. Positive, flat and negative platform-level effects survive even exact parallel untreated trends. These integers are constructed witnesses. The proprietary panel was unavailable.

A low-disclosure summary is possible. If all that is known about the weighted distribution of d is its support $[l, u]$ and mean m, with $l < u$, convexity gives

$$e^{-m} \le H \le \frac{u-m}{u-l}e^{-l} + \frac{m-l}{u-l}e^{-u}.$$

A point mass at m achieves the lower bound. Masses on l, u with mean m achieve the upper bound. These are optimal over the stated distribution class before adding category-specific constraints. Weights by transaction count or by number of categories generally answer a different question.

Which historical comparison can March affect?

Figure III and Appendix D.7 of [1] compare the matched-firm estimate with an estimate obtained by substituting growth-adjusted 2023 outcomes for control firms' post-period outcomes in July–December 2024. The factor for sales uses platform aggregates in January–March 2023 and 2024. March therefore belongs to the calibration window and is the pilot month in Eq. (7). Figure I starts in April, with the first point drawn at zero. Its denominator counts firms and its numerator requires recorded requests. It supplies neither an exact March count nor the sales share of pilot participants. The public records inspected do not identify a March sales effect.

For a ratio of quarterly sums, let w be March's share of the 2023 first-quarter sales denominator. Let g, h denote gross year-over-year growth in January–February and March, respectively. Then $F = (1-w)g + wh$. If h^0 is March growth absent the pilot, and January–February was unaffected, $F - F^0 = w(h - h^0)$. Simply dropping March instead gives $F - g = w(h - g)$, which also captures unrelated seasonal and economic changes.

An exact sensitivity calculation for the matched estimate

Keep sample, regressors and positive regression weights fixed. Let D be the treated-by-post indicator of Eq. (1), and v its weighted residual after projection on the firm-by-cohort and cohort-by-month indicators. Define $h_j = w_j^{\text{reg}} v_j / \sum_k w_k^{\text{reg}} v_k D_k$, assuming nonzero residual variation. The coefficient on D equals $\sum_j h_j Y_j$ by partial regression [3]. For the set C of substituted control observations, $x_j \geq 0$ denotes historical sales. Put $q_j = -h_j$ there. The change between two positive calibration factors is exactly

$$\widehat{\beta}(F_2) - \widehat{\beta}(F_1) = -\sum_{j \in C} q_j \log \frac{1 + F_2 x_j}{1 + F_1 x_j}.$$

Write $t = \log(F_2/F_1) \geq 0$. Each log ratio lies between zero and t. Define $Q_\pm = \sum_{j \in C} \max(\pm q_j, 0)$. The coefficient change is bounded below by $-Q_+ t$ and above by $Q_- t$. This allows mixed regression weights. For all control-post cells $Q_+ - Q_- = 1$, because the control-post indicator equals Post minus D, and Post is in the cohort-month span.

Complete cohort rectangles with separable firm and time weights suffice for $q_j \geq 0$ on control-post observations: the residual within each cohort is $(\text{Treat}_i - \overline{\text{Treat}})(\text{Post}_t - \overline{\text{Post}})$. Replacing only a subset of these cells then gives total mass at most one. In this case raising the calibration factor weakly lowers the sales coefficient, and the absolute change is bounded by $|\log(F_2/F_1)|$. A pilot that inflated the calibrator would make the corrected historical estimate weakly larger. Actual residual weights for Figure III were unavailable for direct inspection.

The blue sales point in Figure III reads about 0.35 on its log scale. To lower that point to zero through recalibration alone, the factor would have to rise by at least about $e^{0.35} - 1 = 42\%$ under the nonnegative-weight condition. A conservative digitization of the marker gives 0.340–0.360 log points, or about 40.5–43.3% for the threshold. The spread reflects graphical resolution. This is a necessary change in the *calibration factor*, not an estimate of the pilot's effect or an inference about statistical significance. The distance between historical and baseline points is much smaller; erasing their difference is a separate target. In a synthetic weighted panel, direct refits reproduced the exact sensitivity formula to within 4×10^{-15} for three recalibrations.

Model assessments and review context. Codex assisted in the derivation, source checks and drafting. Synthetic certificates check the algebra and numerical implementation. External replication on the platform data and independent peer review remain outstanding. Any model assessment is tied to the final PDF.

References. [1] P. J. Barwick, L. Fang, J. Liu, H. Xu and Z. Yuan, AI as Shared Business Infrastructure, NBER WP 36003 (2026), Eqs. (1), (7), Figs. III–IV and App. D.7. [2] B. Callaway, A. Goodman-Bacon and P. H. C. Sant'Anna, Difference-in-Differences with a Continuous Treatment, NBER WP 32117 (2024). [3] R. Frisch and F. V. Waugh, Partial Time Regressions as Compared with Individual Trends, Econometrica 1 (1933), 387–401.