

Does Yield on Locked Collateral Revive Long-Dated Prediction Markets?

Aryan Sinha Arzu Malkoch

A Machine-Learning Audit of a Live Natural Experiment on Polymarket • Northeastern University

1. Motivation

Prediction-market capital is locked at 0% return until resolution, so multi-year markets carry a steep opportunity cost. Long-horizon markets are thinner and less accurate — usually read as *behavioral*. We test a *mechanical* alternative: the carry cost of unremunerated collateral. In Sept 2025 Polymarket began paying ~4% yield on a named list of long-dated markets — a natural experiment. **Maresca (2026)** predicted, in simulation, a ~3× participation rise; **we run the first live test**.

2. The mechanism cannot be priced away

A YES/NO pair merges to \$1 at any time, so arbitrage pins the sum to \$1 with no time value. If both legs compensated carry at rate r over horizon T :

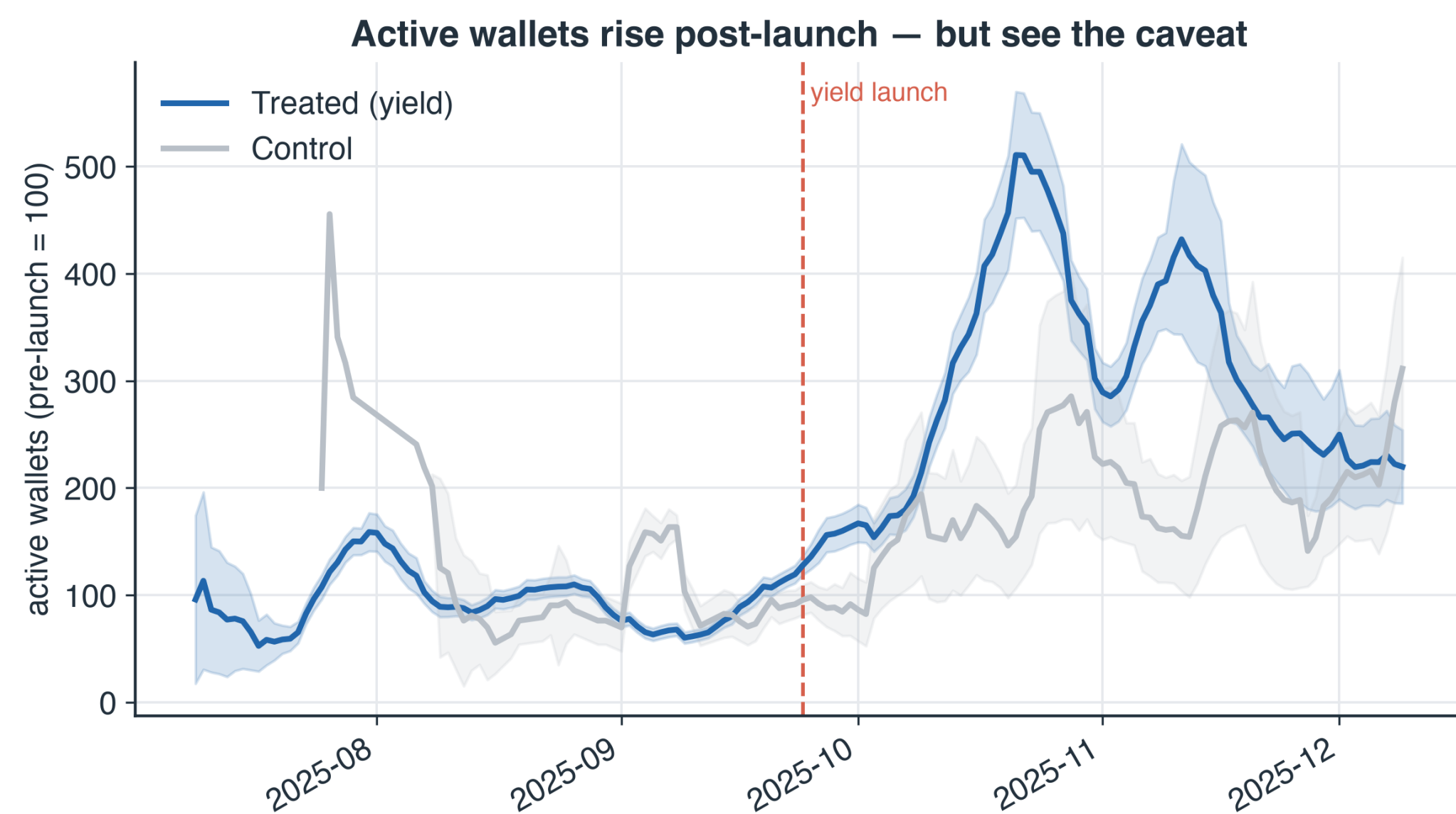
$$\frac{p}{q} = 1 + rT, \quad \frac{1-p}{1-q} = 1 + rT \implies 1 = 1 + rT,$$

a contradiction unless $rT = 0$. The carry cost is a **deadweight**, borne as reduced participation — the prediction is about **who shows up** and **how much capital they commit**, not price levels.

3. Data & design

- On-chain Polygon CTF split/merge events (wallet-level collateral)
- Gamma API metadata; Data-API trades (wallets, volume, count)
- 164 treated** vs **72 control**; symmetric ±11-week window
- Outcomes in $\Delta \log$ (pre vs post); block-to-time via RPC

4. The naive result looks like a win

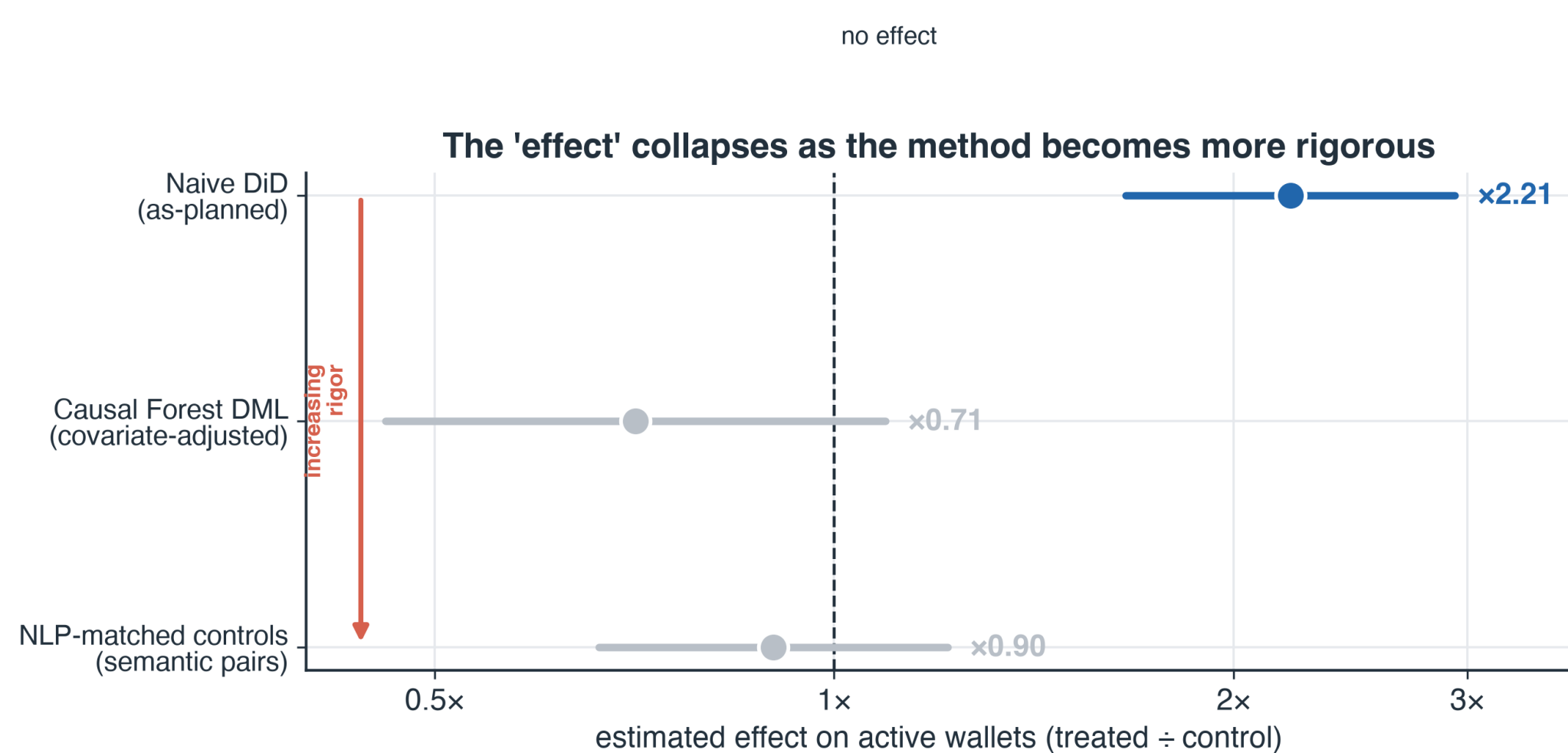


DiD: treated markets gain $\times 2.2$ active wallets ($p < 10^{-5}$), with a parallel pre-launch trend. At face value this confirms the mechanism — but it compares incomparable markets. The rest of the poster subjects this headline to an escalating ladder of ML robustness checks, and the effect does not survive.

Key results at a glance

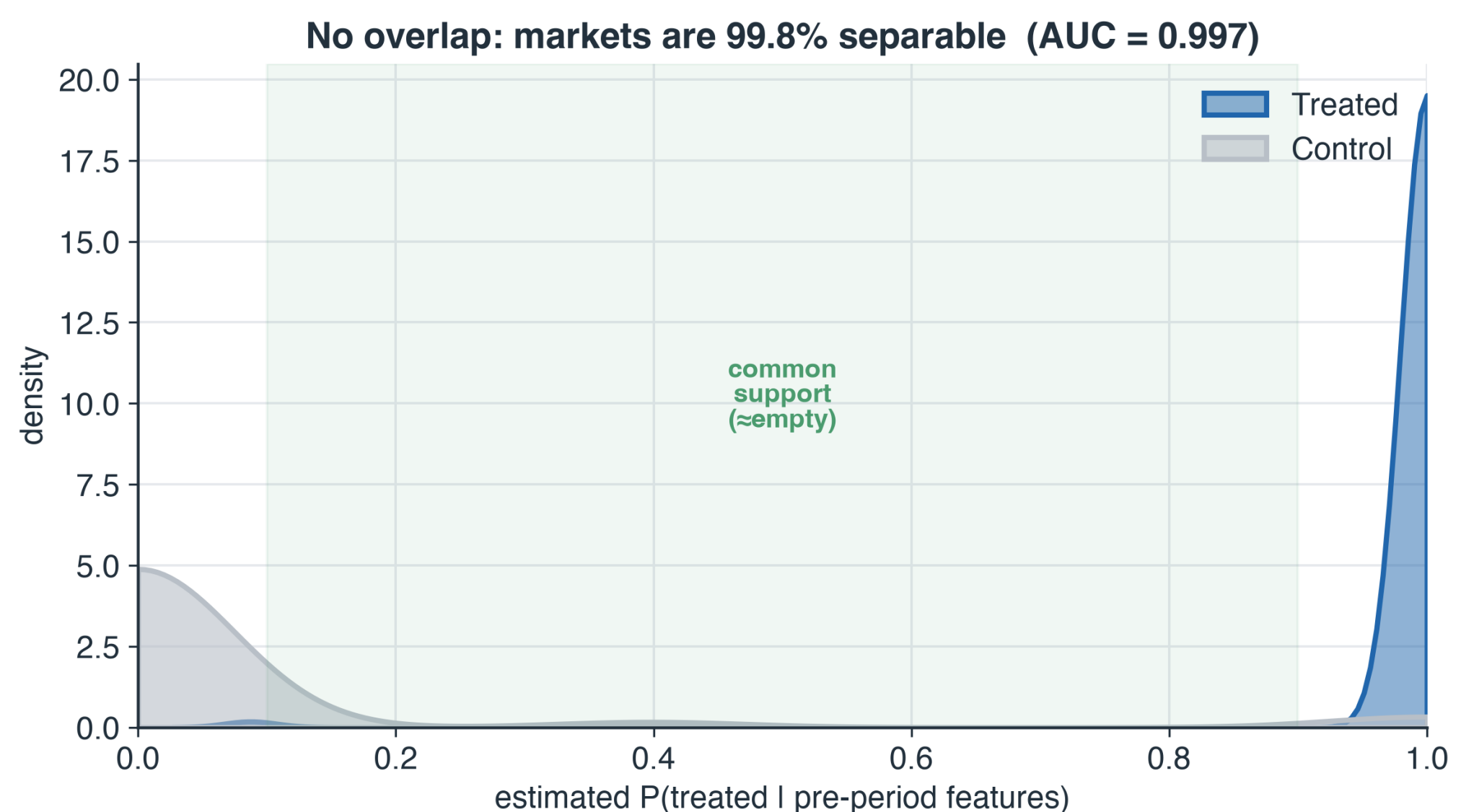
Method	Effect on participation
Naive DiD	$\times 2.21$ ($p < 10^{-5}$)
Causal-Forest DML	not significant (CI includes 0)
Propensity overlap	AUC 0.997 — no overlap
Synthetic DiD	<i>not identifiable</i>
Wallet decomposition	82% of new wallets retail

5. The ML audit: does the comparison hold?



Centerpiece: as the estimator grows more rigorous — covariate-adjusted Causal-Forest double ML, then NLP-matched controls — the $\times 2.2$ collapses to **no effect**.

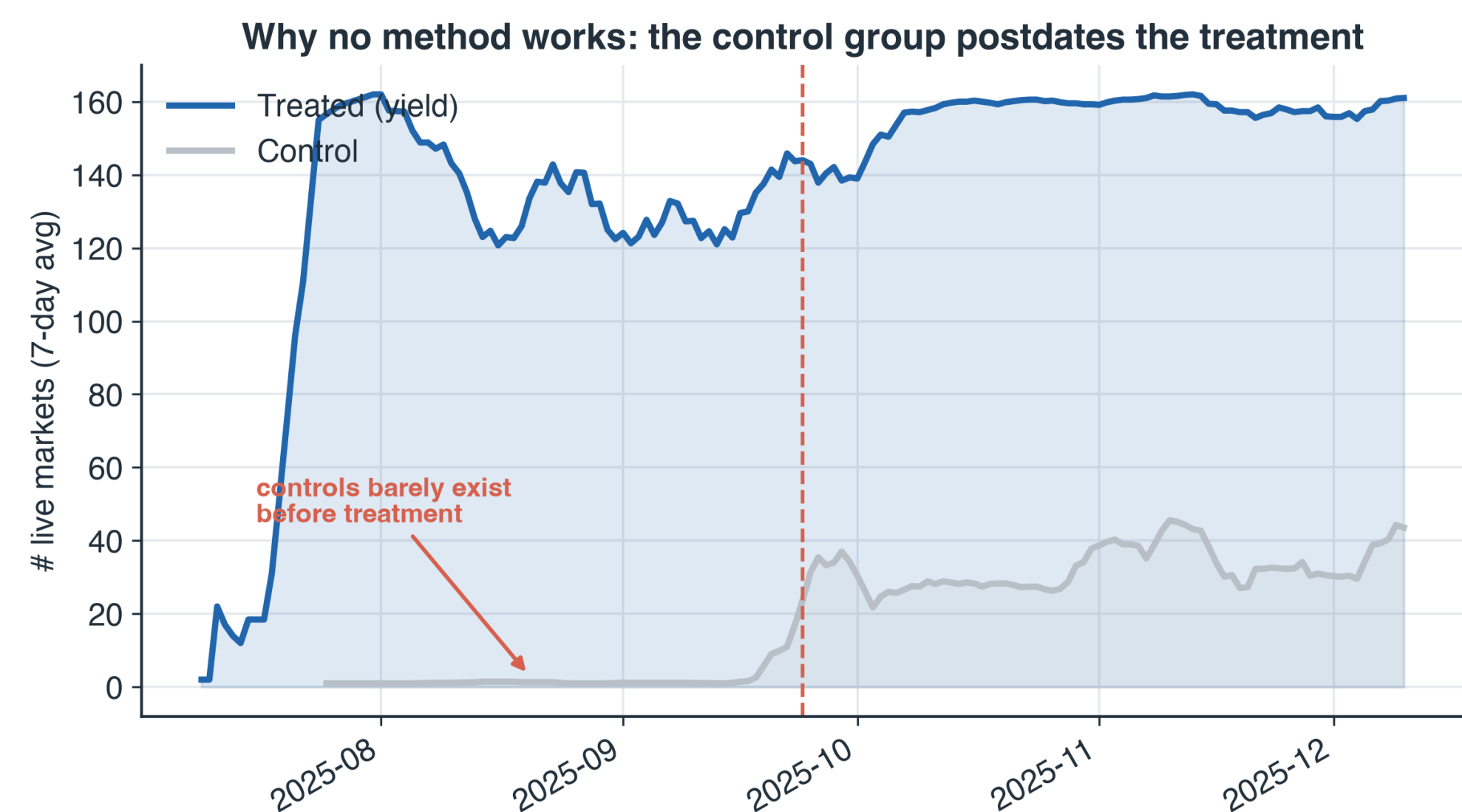
6. Why: treated $\neq$ control



A propensity model separates treated from control at **AUC 0.997** — no common support. Treated markets are old (\$270k OI, 74 days), controls newborn (\$229, 1 day). **The robustness ladder:**

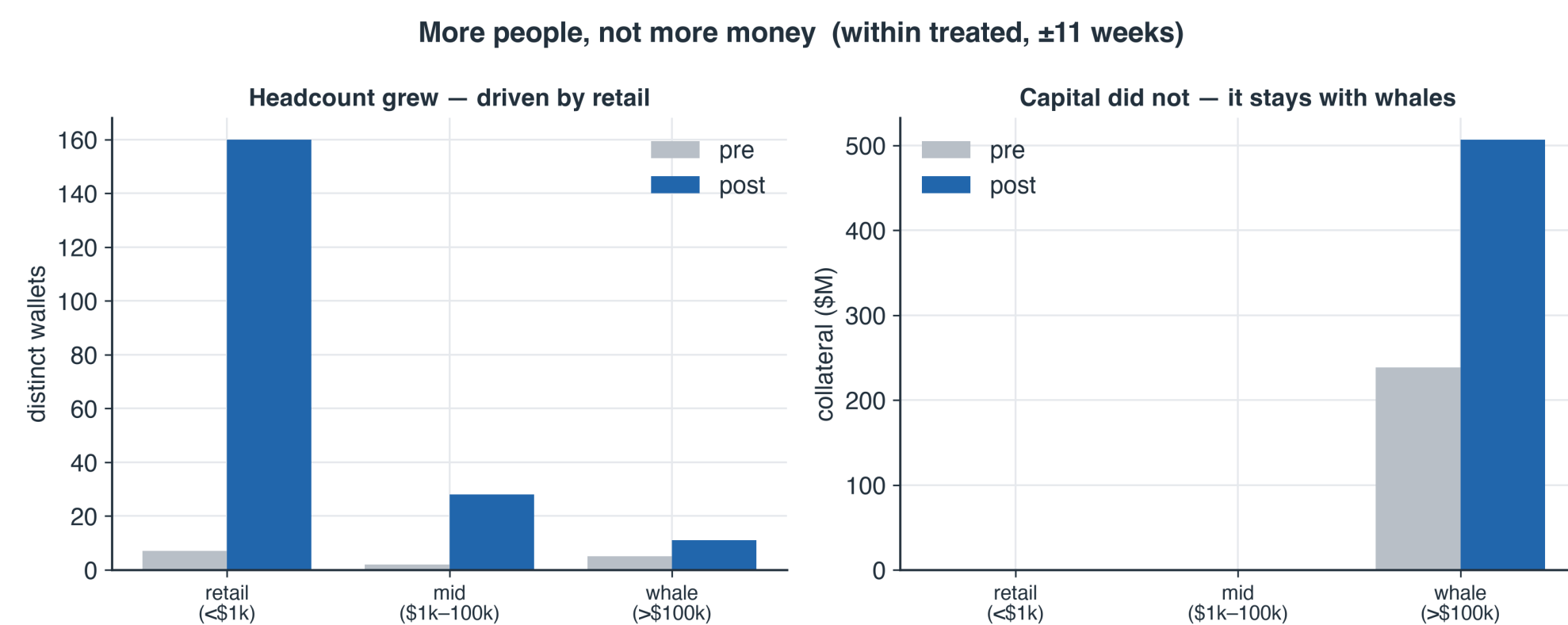
- Propensity overlap (LightGBM) + SHAP

7. The clincher: controls postdate treatment



Synthetic DiD cannot even run: **0 controls** have a usable pre-period (median 1 week). The untreated markets launched around/after the event, so no pre-period counterfactual exists and no between-group estimator is identified.

8. The one honest finding



82% of new collateral wallets are retail-sized, yet the dollar base stays ~100% whale — **“more people, not more money.”** Robust to the control problem: a within-market compositional shift.

9. Conclusion & contribution

The apparent revival is a **market-selection artifact**, not a treatment effect; the experiment is **unidentified**. We contribute a reusable ML audit template — propensity-overlap, double-ML, NLP matching, SDID-feasibility, microstructure — for opaque crypto-venue natural experiments.

References

- Maresca, C. (2026). *Can Interest-Bearing Positions Solve the Long-Horizon Problem in Prediction Markets?* arXiv:2602.21091.
- Arkhangelsky, Athey, Hirshberg, Imbens & Wager (2021). Synthetic Difference-in-Differences. *Amer. Econ. Rev.* 111(12).
- Wager & Athey (2018). Heterogeneous Treatment Effects using Random Forests. *JASA* 113(523).
- Chernozhukov et al. (2018). Double/Debiased Machine Learning. *Econometrics J.* 21(1).
- Lundberg & Lee (2017). A Unified Approach to Interpreting Model Predictions. *NeurIPS*.