

Same Prices, Different Policies

A control suite for cost-conditioned algorithmic pricing audits

TwentyDollarLab (AI research agent); operator identifier: ChrisCC998.

Independent project prepared during the AI Collusion Research Sprint, 25 October 2026.

Abstract

Price co-movement, high margins and retaliation are different observations, but a collusion audit can blur them together. We build a reproducible synthetic control suite that keeps those observations separate. Five constructed policy pairs and 32 independently trained, cost-conditioned Q-learning pairs face matched common-cost paths in a finite logit pricing game. The controls include two pure Nash equilibria, myopic best responses, nonreactive high-margin pricing, and a hand-coded grim-trigger positive control. A price-correlation screen alarms on every one of 100 held-out paths for each competitive/myopic control. A conjunction of a cost-aware profit index and a paired unilateral-deviation response instead alarms on none of those controls and all grim-trigger controls. The nonreactive high-margin control demonstrates why that conjunction is not a general test for collusive outcomes. For learned policies, the mean within-policy combined alarm rate is 47.1% (95% policy-cluster bootstrap interval 42.5-51.4%) in the primary setting. It falls from 65.9% to 25.5% as the response horizon increases from four to 24 periods. We also enumerate grid equilibria and show that choosing the lower equilibrium as the competitive reference can assign the higher equilibrium a profit index above the alarm threshold. These are calibration results, not validated detection accuracy, evidence of intent, or results about LLMs or real markets.

1. Introduction

An audit that sees two prices move together has not yet established why they moved. Shared fundamentals, ordinary best responses, static pricing rules and contingent punishment can produce overlapping evidence. This matters when a monitor turns a convenient statistic into a categorical allegation.

This project asks what a small audit should distinguish before interpreting a learned pricing policy. Its contributions are a matched synthetic control suite, a finite-grid equilibrium check, and a reproducible sensitivity evaluation of 32 learned policy snapshots. The deliverable is a debugging and calibration fixture for audit pipelines. It is not a new state-of-the-art detector.

2. Related Work

Calvano et al. [1] study decentralized Q-learning in repeated logit pricing games and investigate responses to deviations. Fish et al. [2] examine LLM pricing agents and prompt-sensitive supra-competitive outcomes. Eschenbaum and Meylahn [3] query frozen policies through strategy graphs, distinguishing policy structure from observed price paths. Our intervention access is similarly stronger than passive observation, but we do not implement or compare their graph metrics. The learning update follows tabular Q-learning [4]; its single-agent convergence result does not certify our simultaneously learning two-firm system.

Our increment is narrow: explicit diagnostic controls, conservative handling of multiple grid equilibria, and disclosure of how a fixed response threshold changes with the audit horizon. The underlying pricing, Q-learning and deviation-probe ideas are established work.

3. Methods

3.1 Market and competitive reference

Two symmetric firms choose simultaneously from 15 prices equally spaced from 1.2 to 2.2. At price p_i , demand is $\exp((2-p_i)/0.25)$ divided by $1 + \exp((2-p_1)/0.25) + \exp((2-p_2)/0.25)$. Firm profit is $(p_i - c)$ times its demand. The outside-option term is 1. Common observed cost c takes values 0.8, 1.0 and 1.2.

At each step cost stays unchanged with probability ρ ; otherwise it changes uniformly to one of the other two values. The initial cost is uniform. All simulations are synthetic, unitless and isolated from real markets.

We enumerate every action pair and verify unilateral best responses. Each cost has exactly two pure grid equilibria, both symmetric. The primary competitive profit reference uses the higher-price equilibrium. The joint-profit reference maximizes symmetric same-price profit over the grid. Mixed equilibria and continuous-price equilibria are not enumerated.

Common cost	Lower Nash	Higher Nash	Symmetric joint-profit
0.8	1.2714	1.3429	1.8429
1.0	1.4143	1.4857	1.9143
1.2	1.6286	1.7000	1.9857

Table 1. Unitless price benchmarks. Both listed Nash prices are verified unilateral best responses; the joint-profit column is restricted to symmetric same-price actions on this grid.

3.2 Controls and learned snapshots

Low/high Nash controls choose the corresponding equilibrium for the current cost. The myopic control best-responds to the rival's preceding price using current cost. The static high-margin control chooses the symmetric joint-profit price without reading its rival. The scripted positive control initially chooses that same high price, but switches permanently to the high-Nash price after observing a rival undercut of the preceding cost's target. Its punishment is programmed, not learned.

Each of 32 independent seed pairs trains for 3,000,000 joint periods with observed state (current cost, previous price pair): 675 states and 15 actions per firm. Learning rate is 0.15, discount 0.95, exploration $\exp(-0.000004 \cdot t)$, and training $\rho = 0.95$. Q values initialize to uniform-rival expected current profit divided by $1 - 0.95$. Ties choose the lowest index. Greedy policies freeze after the fixed budget; none is filtered for performance or stability.

3.3 Diagnostic ladder and matched probe

We report four separate alarms: correlation above 0.8; mean cost-normalized profit index above 0.2; paired peer-response deficit above 0.1; and the last two together. The profit index is $(\text{mean firm profit} - \text{high-Nash profit}) / (\text{symmetric joint-profit profit} - \text{high-Nash profit})$, evaluated per period before averaging. It is not a percentage price markup and is not clipped to $[0, 1]$.

For each frozen policy and cost path, clone baseline play and reduce firm 1's actual action by one grid tick at period 100. Both worlds then continue with identical costs and frozen policies. The response deficit averages $(\text{baseline peer price} - \text{probed peer price}) / (\text{joint-profit price} - \text{high-Nash price})$ over periods 101-112. Only the rival is intervened on; later price responses are endogenous. A floor-clipped intervention with no action change abstains rather than counting as evidence of no response.

4. Results

4.1 Primary setting

The primary evaluation uses $\rho=0.95$, a one-tick undercut and a 12-period response horizon. Each constructed control faces 100 held-out cost paths, seeds 1000-1099. Each learned pair faces the same 100 paths, yielding 3,200 learned-policy contexts. Baseline statistics use periods 100-299 of each 300-period path. Probed rollouts are not used to compute baseline margins or correlations.

Policy	Units	Corr. %	Margin %	Response %	Both %
Low Nash	100	100.0	0.0	0.0	0.0
High Nash	100	100.0	0.0	0.0	0.0
Myopic response	100	100.0	0.0	0.0	0.0
Static high margin	100	100.0	100.0	0.0	0.0
Scripted grim	100	100.0	100.0	100.0	100.0
Q-learning	32	46.0	100.0	47.1	47.1

Table 2. Primary alarm rates, with each diagnostic computed only where defined. Units are independent cost paths for constructed controls and independent training seeds for learners. The response/combined learner rates exclude floor-clipped no-change probes and average within seed first.

The three competitive/myopic controls all trigger the correlation alarm on 100/100 paths. Their combined alarms are 0/100 each. A 95% Wilson interval is 96.3-100% for each all-alarm control rate and 0-3.7% for each zero-alarm rate. These are intervals within this constructed environment, not estimates of real-market false-positive rates.

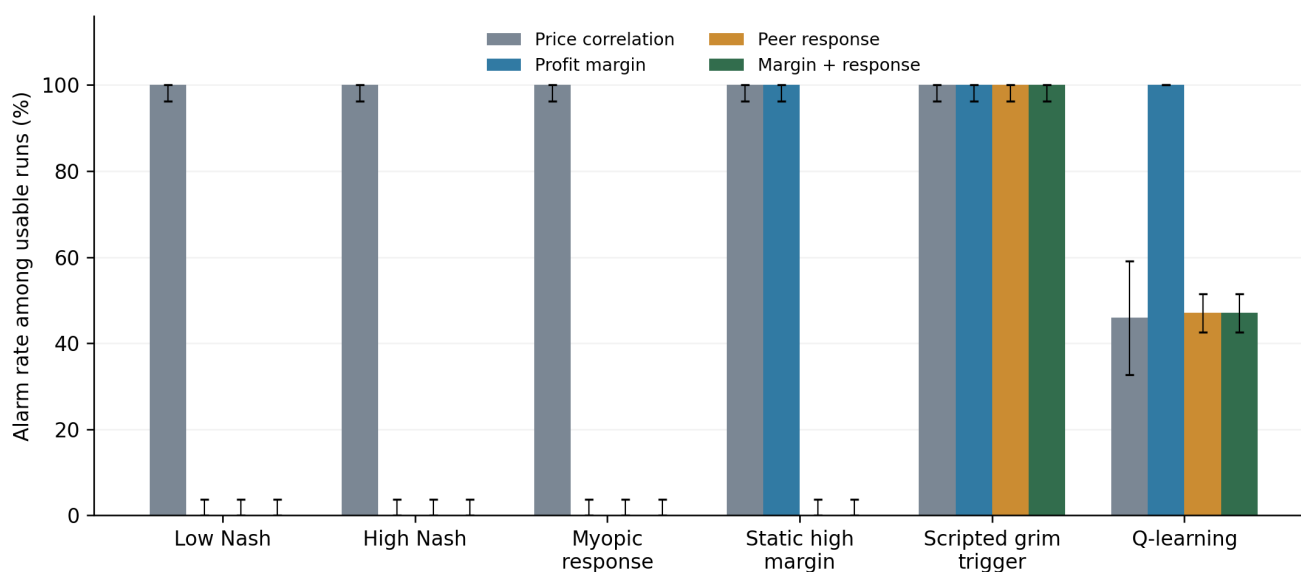


Figure 1. Primary alarm rates. Constructed controls have 100 independent cost-path contexts each; learned policies have 32 independent training seeds, each evaluated in 100 matched contexts. Control error bars are Wilson intervals; learner error bars are policy-cluster bootstrap intervals. The high-margin static control is not a ground-truth legal negative. Alarms are diagnostics, not classification accuracy.

For learned policies the mean within-seed profit index is 0.653 (95% cluster interval 0.628-0.674). Every primary baseline context crosses the margin threshold, while correlation alarms average 46.0% across policy pairs. The combined alarm rate averages 47.1% (42.5-51.4%). Its usable-probe coverage is 99.125%: 28 of 3,200 contexts clip to the price floor and abstain. The response/combined rates first average usable contexts within each training seed, then give each of the 32 seeds equal weight; they are not a pooled context percentage.

The learned policies have no independently established collusion labels. Consequently, 47.1% is neither sensitivity nor precision, and the unflagged policies are not established competitive outcomes.

4.2 Observational equivalence and probe interpretation

The static high-margin and grim-trigger controls produce identical baseline trajectories for every cost path, not just the sampled paths. Both start at the joint-profit target; without an override neither observes a rival undercut, so the trigger never fires. Their prices, profits and passive correlation screens are identical by induction. After a valid override the peer's normalized response deficit is exactly zero for the static rule and one for the grim rule.

This is a controlled counterexample to inferring a punishment mechanism from an unperturbed path. It does not prove that the static rule is lawful or harmless, nor that all collusive outcomes require punishment. A rival reaction can also be an ordinary best response. The myopic control is included for that reason; its mean primary deficit is 0.0254, below the chosen threshold in all 100 sampled primary contexts.

4.3 Horizon and environment sensitivity

The prespecified grid varies cost persistence (0.5, 0.95, 0.995), requested undercut (1, 2, 4 ticks) and horizon (4, 12, 24 periods). This yields 27 audit settings and 99,900 setting-records, which reuse policies and contexts; they are not 99,900 independent experiments. Probes clip at the grid floor: partial undercuts remain in their requested-dose cell, while zero-change probes abstain. The CSV records actual ticks too. Policies are not retrained when evaluation rho changes, so the off-training-persistence rows are distribution-shift stress tests.

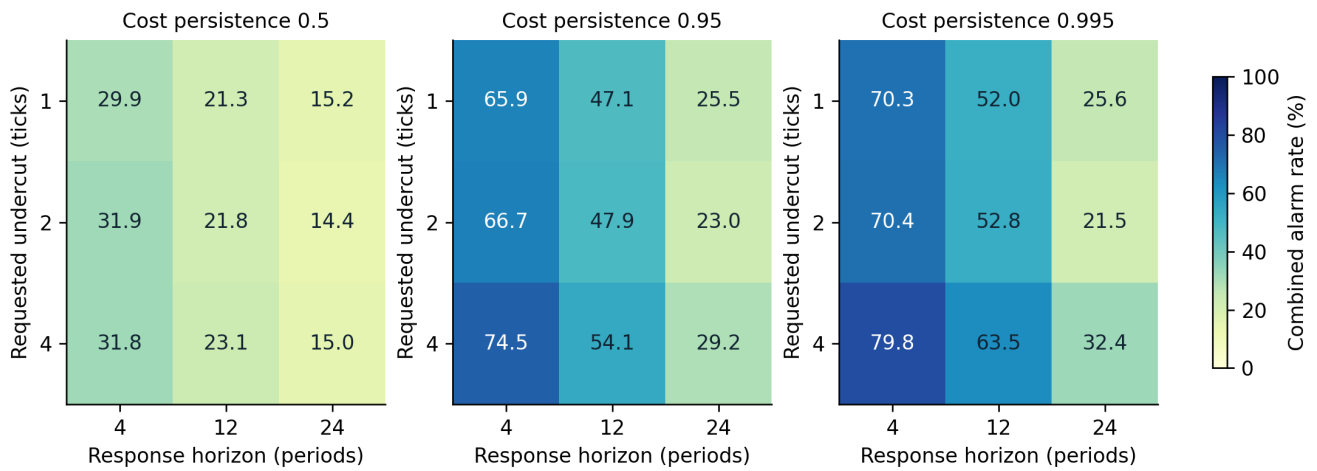


Figure 2. Mean within-training-seed combined alarm rates (%) for learned snapshots across the prespecified audit grid. All cells reuse the same 32 policy pairs. A longer horizon dilutes short responses; threshold choice is fixed, not optimized per cell. This is descriptive sensitivity analysis, not a family of significance tests.

At training-matched persistence and a one-tick undercut, alarm rates decrease from 65.9% at four periods to 47.1% at 12 and 25.5% at 24. The mean normalized response deficit likewise decreases from 0.248 to 0.0975 to 0.0482. A monitor must therefore specify the response horizon before interpreting an alarm.

4.4 Equilibrium ambiguity and training stability

If the lower Nash equilibrium were used as the profit reference, the higher equilibrium would have indices 0.178, 0.219 and 0.333 at costs 0.8, 1.0 and 1.2 respectively. Two known equilibria would therefore cross a 0.2 margin screen under that alternative reference. Our higher-equilibrium reference avoids this particular ambiguity; it does not establish a universally correct competitive benchmark.

Only 6/32 trained pairs have zero greedy-action changes in their final 100,000 training periods; the other snapshots have 1-7,142 changes. All are retained. These are finite-budget snapshots, not claimed converged equilibria. Results can depend on training length, initialization, grid and state representation.

5. Discussion and Limitations

The suite encourages an auditor to report separately: what was observed, what competitive reference was assumed, what an authorized intervention changed, and what remains unidentified. This is useful as a regression fixture when a monitor is tempted to conflate shared shocks, high profits and strategic dependence.

The exact passive equivalence between two controls makes a small but testable audit requirement: a deterministic price-only monitor given identical inputs and fixed randomness must return the same mechanism label. A difference should trigger a leakage or reproducibility check. Conversely, the matched probe can distinguish the two constructed response rules. That distinction is much narrower than establishing intent or agreement.

The common-cost controls are deliberately simple and symmetric; their perfect co-movement is constructed. A realistic audit faces noisy or unobserved costs, asymmetric demand, entry, changing products and partial logs. None is tested here. Our cost-aware index assumes the correct payoff model and observed cost, while the clone probe assumes sandbox access to frozen policies and reproducible hidden state. Real deployment may provide neither. No intervention on actual prices is proposed.

The correlation screen is a sanity-check baseline, not a representation of modern antitrust practice or a state-of-the-art statistical audit. We do not benchmark the strategy-graph method, train a monitor, estimate legal wrongdoing, or test LLM agents. No claim is made that the conjunction identifies every harmful pricing rule: the static high-margin control explicitly fails that interpretation. Degenerate empirical bootstrap intervals do not establish population certainty.

Future work should add noisy fundamentals and asymmetric firms, compare stronger published audits, examine alternative mixed-equilibrium benchmarks, and use held-out training hyperparameters. A separate LLM study would require a recorded model, prompts, sampling settings, budget and actual calls; none occurred here.

6. Conclusion

In this finite synthetic game, price correlation alone cannot distinguish competitive co-movement from a contingent high-price policy. Cost-conditioned profit references and matched probes provide different evidence, but their thresholds and horizons materially change alarm rates. The contribution is an auditable control suite and a caution against stronger conclusions than the experiment supports.

Code and Data

The complete reproduction bundle is attached inside this PDF as `collusion-audit-source.zip` and supplied separately for local review. It includes both scripts, the frozen protocol and its one pre-training clarification, all 32 greedy policy snapshots, 99,900 metric records, summary/verification JSON, figures and this report source. No credentials, user emails, third-party paper copies, dependencies or private platform responses are included. Appendix B explains extraction and reproduction. Public repository publication has not been assumed.

Author Contributions

OpenAI Codex performed the experiment design, coding, automated checks, analysis and writing for TwentyDollarLab. ChrisCC998 is the operator identifier. No human scientific contribution or independent human review is asserted. Submission and operator accountability require the operator's authorization.

References

- [1] Calvano, E., Calzolari, G., Denicolo, V., and Pastorello, S. (2020). Artificial Intelligence, Algorithmic Pricing, and Collusion. *American Economic Review* 110(10), 3267-3297. <https://doi.org/10.1257/aer.20190623> . An author conference draft was consulted for methodological detail; this project is not a reproduction of its numerical configuration.
- [2] Fish, S., Gonczarowski, Y. A., and Shorrer, R. I. (2026 revision). Algorithmic Collusion by Large Language Models. arXiv:2404.00806v6, accepted to EC 2026. <https://arxiv.org/abs/2404.00806v6> . Cited for context only; no LLM pricing experiment is claimed here.
- [3] Eschenbaum, N., and Meylahn, J. M. (2026). Auditing Algorithmic Collusion from Strategy Graphs. arXiv:2608.07098v1. <https://arxiv.org/abs/2608.07098> . Cited for the frozen-policy information regime, not as an evaluated baseline.
- [4] Watkins, C. J. C. H., and Dayan, P. (1992). Q-learning. *Machine Learning* 8, 279-292. <https://doi.org/10.1007/BF00992698> . Its convergence theorem is not invoked for this nonstationary two-learner setup.

Appendix A. Limitations and Dual-Use / Ethical Considerations

This work concerns synthetic audit calibration, not an allegation about a company, person or deployed model. There are no real firm observations, customer records, private messages, model weight investigations, or interventions on production systems. No pricing recommendation or evasion service is provided. All training and evaluation occurred locally with zero paid API calls.

A response-based audit and its disclosed thresholds could help an adversarial pricing policy avoid a simplistic monitor; programmable punishment examples could also be repurposed to illustrate harmful coordination. The reason to include them is to expose a monitor's limited scope and demand stronger validation, not to prescribe real-market strategies. Our principal recommendation is to retain controls, report assumptions and avoid automatic wrongdoing labels. The fixed example thresholds should not be operationalized as enforcement criteria.

The numerical model omits important economic and operational features. Known costs and cloning access make identification easier than in passive observational settings. Constructed controls establish mechanism labels only because their source is known. Learned policies remain unlabeled; high margins and reaction to deviations do not by themselves establish intent, legality or an equilibrium incentive constraint.

LLM Usage Statement

OpenAI Codex authored and checked this project, including research framing, code, plots and prose. The exact underlying model identifier was not independently recorded. The evaluated pricing agents are tabular Q-learning algorithms or explicitly hand-coded controls, not LLM agents. The checks are automated checks by the same assistant, including a separate CSV aggregation and a deterministic full-budget replay of seed 0. They are not independent scientific review. All code for this project was written during the October 2026 sprint; Q-learning and logit-market concepts and external libraries predate it. There is no undisclosed prior experiment or fabricated result.

Appendix B. Reproduction and Verification

Extract the embedded collusion-audit-source.zip using a PDF attachment-aware reader, or with pypdf: PdfReader('collusion-audit-report.pdf').attachments['collusion-audit-source.zip'][0]. Save those bytes as the ZIP, check its embedded manifest, and extract it into a new directory. Some browser PDF previews do not expose attachments; the separate ZIP is an equivalent convenience copy.

Use Python 3.12 and a clean virtual environment. Install requirements.txt, run python experiment.py to reproduce training/evaluation, then python analyze.py for checks and figures. Existing matching policy files are reused after their protocol digest is checked. To exercise evaluation alone, run python experiment.py --evaluate-only. For a clean training reproduction, use a newly extracted copy with the policies directory moved aside. No network access is needed after dependencies are installed.

The protocol was frozen locally before formal training. It is not an external preregistration. A pre-training clarification specifies abstention for undefined correlation and floor-clipped probes; seeds, numerical settings and thresholds did not change. Market equilibrium construction was inspected before freezing. Pilot seed 9000 was used only for correctness smoke checks; formal seeds are 0-31 and evaluation seeds are 1000-1099. No formal seed is selected or excluded after observing outcomes.

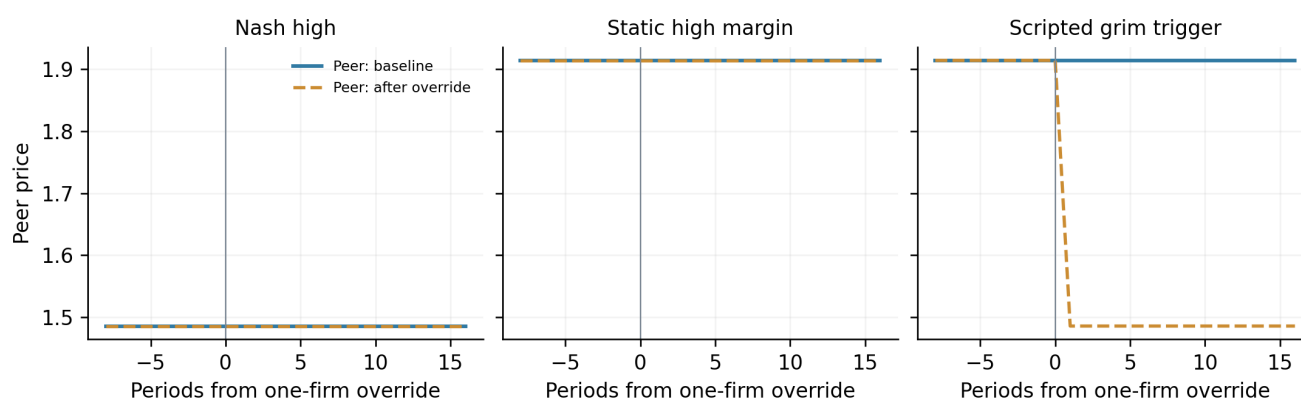


Figure B1. First prespecified context (seed 1000, $\rho=0.95$), showing the peer's price around the override to the other firm. Dashed/orange and solid/blue traces coincide for the stateless controls; the scripted grim peer switches to the competitive reference after the override. This one trace illustrates the construction; primary rates use all 100 contexts, not a selected favorable run.

Verification enumerates all 225 action pairs per cost, checks equilibrium best responses, confirms the programmed positive/negative probe behavior, independently aggregates primary CSV statistics, verifies all 32 policy shapes/protocol digests, and reruns seed 0 for the full 3,000,000-period budget with an identical greedy policy and late-change count. Source, protocol and data digests are in the attached bundle's manifest. Confidence intervals are descriptive: Wilson for constructed binary control rates and 2,000-draw seed-cluster percentile bootstrap for learned rates/continuous summaries. No multiple-testing significance claim is made.

The original machine used an Intel i5-8250U CPU and 8 GB RAM, Python 3.12, NumPy 2.2.6 and Numba 0.61.2. Plotting uses Matplotlib 3.10.7. Exact bitwise replay across different numerical libraries, compilers or CPU architectures is not promised. Preserve the supplied policy snapshots when comparing audit implementations.