Contract-as-Instrument Breaks Under Selection: Safe Randomization and Online IV Learning with Endogenous Participation

Liz Lemma Future Detective

January 16, 2026

Abstract

Platforms increasingly set incentive weights algorithmically while facing endogenous worker participation (opt-in/opt-out, multi-homing) and noisy performance telemetry. Building on recent work that treats the posted contract as an instrument to correct measurement error in multitask contracting, we show that endogenous participation fundamentally breaks the contract-as-instrument logic: when unobserved agent quality affects outcomes and is correlated with participation, contracts shift the composition of observed agents, inducing selection bias and invalidating naive moment conditions. We propose a minimal 'safe experimentation' design—post-participation mean-zero randomization of contract slopes (or quasi-exogenous audits)—that restores instrumental validity without requiring parametric selection models. We characterize identification via conditional moment restrictions, give finite-sample error bounds for IV/GMM with selected samples, and provide an online learning algorithm with sublinear regret. The results provide a tractable blueprint for learning interpretable linear incentive rules in modern platforms where participation and composition effects are first-order.

Table of Contents

4. 4. Failure of contract-as-IV under selection: a two-type example and general bias decomposition; when the source paper's moment conditions fail.

5. 5. Safe instruments: (A) post-participation slope randomization; (B) randomized audits/repeated measurements; formal exogeneity and relevance conditions; identification theorem.

6. 6. Estimation: IV/GMM estimator on selected samples; finite-sample bounds; weak-instrument diagnostics; choosing randomization variance (precision–welfare tradeoff).

7. 7. Online learning: algorithm (estimate θ^*, update baseline b_t); regret analysis with self-normalized martingale tools under selection; participation feasibility constraints.

8. 8. Extensions (structured, optional): observed covariates and group fairness; multi-homing with platform competition (reduced-form); delayed outcomes; nonstationarity.

9. 9. Discussion: implementation in platforms; audit design; policy implications; limitations and open problems.

1 Introduction

Digital platforms routinely pay agents—drivers, sellers, creators, curators, moderators—through contracts that are both *multitask* and *adaptive*. Multitask because compensation is tied to a vector of measurable signals (acceptance rates, response times, content formats, customer ratings, delivery completion, and so on), and adaptive because the platform continuously updates these incentives as it learns which behaviors generate downstream value. A central practical complication is that the data the platform uses to learn are themselves endogenously generated: when incentives change, not only do participating agents adjust effort, but the *set of participating agents changes*. The platform therefore faces a selection problem that is inseparable from the incentive problem. Our objective is to formalize this interaction and to isolate a simple design principle that restores valid learning while preserving the operational attractiveness of linear, signal-based pay.

The starting intuition is straightforward. Suppose the platform raises the slope on some measured activity (say, "on-time delivery"), and then estimates the activity's causal contribution to platform value using the observed relationship between outcomes and measured activity under the new contract. If the higher slope disproportionately attracts agents who are intrinsically more productive, better informed, or better matched to the platform (captured in the model by an unobserved shifter), then outcomes rise even holding effort fixed. A naive estimator will attribute this composition-driven improvement to the measured activity itself, overstating its value and leading the platform to further amplify the wrong incentive. Importantly, this failure persists even if the platform is otherwise careful and uses the posted contract as an instrument for activity. The reason is that the contract shifts participation, so the sample in which outcomes are observed is *missing not at random*: the distribution of unobserved determinants of outcomes changes with the instrument. In short, the very lever used to induce variation for identification also moves the selection mechanism, invalidating standard IV logic when applied to the observed (participating) sample.

This selection problem is not merely a technical nuisance; it is a first-order design constraint for online contracting. Many platforms can observe rich activity telemetry only after an agent opts in, logs on, or accepts a task, and they observe downstream value only for those who complete the activity. As a consequence, learning is inherently "on-policy" and filtered through participation. When participation is sensitive to incentives—as it typically is in flexible labor markets and creator ecosystems—a platform that ignores selection can be systematically misled even in large samples. Moreover, the direction of the bias is economically meaningful: if higher-powered incentives attract higher-quality agents, the platform may overpay for measured activity; if they attract lower-quality agents (for instance, because the contract appeals most to those with weaker outside options), the platform may under-

incentivize valuable behavior. Either way, the platform confounds *behavioral responses* with *composition responses*.

We show that this confounding admits a clean remedy that is implementable within the same linear-contract architecture. The key idea is to separate the contract component that drives participation from the contract component used for identification. Concretely, the platform posts a baseline contract that agents can evaluate when deciding whether to participate, and then introduces a mean-zero randomized perturbation to the contract *after* participation is chosen. Because the perturbation is realized only after the selection decision, it does not affect who enters; yet, once realized, it moves incentives and therefore generates exogenous variation in effort and signals among participants. This "post-participation randomization" restores orthogonality between the instrument and unobserved outcome shifters within the selected sample, yielding valid moment conditions and consistent estimation of the platform's underlying task values. In operational terms, the perturbation can be interpreted as a surprise bonus schedule revealed only after opt-in, a randomized weighting of score components, an audit-like adjustment that conditions on realized actions, or a randomized "experiment flag" that changes the mapping from telemetry to pay while keeping the posted baseline unchanged in expectation.

Why does timing matter so much? If randomization is applied before participation, then participation becomes a function of the realized contract, and conditioning on observing outcomes (which requires participation) induces correlation between the randomized contract and unobserved determinants of outcomes. This is the familiar collider logic of selection: even independent shocks become dependent once we condition on an event that both shocks influence. By postponing the random draw until after participation, we break this collider channel. The platform still conditions on participation because it only observes outcomes for participants, but participation no longer depends on the realized perturbation. The perturbation therefore remains independent of the unobserved shifter within the participating sample, and IV/GMM moment conditions become valid *conditional on participation*. The resulting identification requirement is the usual rank condition, but evaluated on the selected sample: the perturbation must generate sufficiently rich variation in observed signals among participants.

Our analysis emphasizes that the platform's experimentation problem is inseparable from its revenue problem. Randomized perturbations are not free: they distort incentives away from the profit-maximizing baseline and may reduce contemporaneous surplus. This creates a precision–revenue tradeoff that the platform must manage over time. The model helps clarify how this tradeoff depends on the strength and geometry of the randomization (captured by its covariance) and on the endogenous participation rate. Intuitively, more variance strengthens the instrument and accelerates learning, but it also increases short-run distortion; similarly, higher participation

increases the effective sample size and improves learning speed, but participation itself is influenced by the baseline contract and by agents' outside options. These forces are familiar in experimentation, but the selection channel changes their relative importance: a platform that chases short-run participation through aggressive baselines may inadvertently weaken identification if participation becomes concentrated among types with low responsiveness along certain task dimensions.

Contributions. We make four main contributions. First, we formalize the selection bias that arises when a platform uses its posted contract (baseline or realized) as an instrument for measured activity in the presence of endogenous participation. Even when measurement error in the activity signal is classical and effort is chosen optimally given the contract, the IV moment generally fails because unobserved outcome shifters vary with the contract through composition. This result clarifies a conceptual pitfall in interpreting contract variation as quasi-experimental variation in effort when participation is flexible.

Second, we provide a simple sufficient condition for selection-robust identification: a mean-zero contract perturbation that is realized *after* participation and is independent of agent types and outcome noises. Under this condition, the perturbation is orthogonal to the residual in the platform's outcome equation within the participating sample, yielding a valid conditional moment restriction. Identification then follows from a selected-sample relevance (rank) condition requiring that the perturbation move observed signals among participants in a nonsingular way.

Third, we translate this identification logic into an estimator and finite-sample learning guarantees. Using the participating rounds only, the platform can run a standard linear IV/GMM regression with the post-participation perturbation as the instrument. Under subgaussian noise and mild regularity ensuring the instrument is not weak on the selected sample, we obtain an error bound that mirrors classical IV rates but with an effective sample size equal to the number of participants. This explicitly quantifies how participation scarcity slows learning even when selection bias is eliminated.

Fourth, we connect learning to online contract design. We study a policy that updates baseline incentives as a function of the evolving IV estimate while maintaining persistent post-participation randomization to preserve identification. Under local curvature conditions on the platform's objective and a participation rate bounded away from zero, we obtain sublinear regret relative to the best fixed baseline contract in hindsight. The result highlights a general message: in environments with endogenous participation, *safe exploration* is not merely about limiting incentive distortion; it is also about structuring randomness so that it survives the selection filter.

Design principles and interpretation. The post-participation perturbation suggests concrete guidance for practitioners. First, *commitment* is essential: agents must know the perturbation distribution ex ante (so participation decisions are well-defined) even though the realization is revealed only after opt-in. Second, *timing* is the source of robustness: to insulate identification from selection, the experiment must not affect the participation margin. Third, *span* matters: perturbations must vary incentives across tasks in a way that generates full-rank covariance in the induced signal variation among participants; otherwise, some task values remain weakly identified. Fourth, *calibration* matters: the scale of randomization should decline with accumulated information if short-run distortion is costly, yet it must remain large enough to avoid weak-instrument problems, especially when participation is low or concentrated. These principles also connect to audit regimes: one can view the perturbation as a randomized "audit weight" applied after participation, which preserves fairness in expectation while still generating quasi-experimental variation for learning.

Limitations and scope. We deliberately work within a stylized but widely used linear-contract framework. The analysis abstracts from risk aversion and from dynamic participation decisions beyond the round-by-round entry choice; in practice, surprise randomization can interact with perceived volatility, trust, and long-run platform–agent relationships. We also treat the perturbation as observable and enforceable, whereas real implementations may be constrained by regulation, transparency requirements, or strategic manipulation of telemetry. Finally, our identification relies on a form of independence between the perturbation and unobservables; if the platform targets randomization based on rich observables, one must verify that the targeting does not reintroduce selection-like dependence. We view these as important directions for extension, but they do not undermine the central lesson: when outcomes are observed only after endogenous participation, the platform must engineer variation that is realized *inside* the selected sample rather than before selection occurs.

The remainder of the paper develops these points formally and situates them within the broader literature on multitask incentives, econometric identification with selection, and online experimentation in platforms and marketplaces.

2 Related literature

Our setting sits at the intersection of three canonical themes: multitask incentives under linear contracts, econometric identification with noisy proxies and endogenous sampling, and online learning when feedback is censored by participation. The common thread is that a platform wants to infer the

value of behaviors it can measure only imperfectly, using data generated by agents who can opt out. We briefly situate our approach relative to these literatures and clarify what is standard versus what is specific to the participation-timing issue emphasized here.

Multitask principal–agent models and linear contracting. The economic motivation for tying pay to multiple observable signals originates in the multitask agency literature, where agents allocate effort across dimensions that differ in social value and measurability. Classical analyses emphasize that when the principal observes only imperfect performance measures, optimal incentives generally distort effort toward measurable tasks and away from hard-to-measure but valuable activities; see, among many others, **?** and the broader discussions in **?**. A key analytical convenience in this literature is the linearity of contracts in observed performance signals, which can be justified either as a tractable approximation or via stronger assumptions such as CARA preferences with normal uncertainty (**?**). Our model adopts the linear-contract architecture, not as a claim of literal optimality, but because it reflects common platform practice and because it makes transparent how incentive slopes map into effort choices through first-order conditions.

Relative to the classic principal–agent benchmark, the central complication in platform contexts is that *the principal does not observe the relevant signals for non-participants*. Traditional agency models typically take the agent as given (or treat participation as a static constraint) and focus on moral hazard and risk-sharing. By contrast, in gig-economy and creator settings, participation is an active margin: agents can log off, multi-home, or select into tasks. This adds a composition channel absent from a fixed-agent analysis: changing the contract changes who shows up, and those compositional shifts can be first-order for both profits and inference. Participation constraints do appear in agency theory, but they are often imposed as a static individual rationality condition pinned down at the optimum. Here, participation is repeatedly realized and interacts with experimentation: the platform learns from those who participate, and the contract both induces effort and filters the sample.

Measurement error, proxies for effort, and IV/GMM. A second strand of related work concerns identification when the econometrician observes a noisy proxy for the endogenous choice variable. In our environment, effort is the latent object of interest, while the platform observes a signal that is an unbiased but noisy measure of effort. This resembles the classical measurement-error problem in linear models, where naive regressions are attenuated and the use of instruments or multiple measurements restores identification; see, e.g., **?**. In platform applications, telemetry is often precisely of this form: it is abundant and high-dimensional, but it is not a

perfect measure of the behavioral construct that enters the platform's value function.

Instrumental-variables and GMM methods provide the standard toolkit for learning causal effects in the presence of endogeneity and measurement noise. Our technical statements mirror familiar IV logic: one needs an instrument that shifts the endogenous regressor (relevance) while remaining orthogonal to the unobserved determinants of outcomes (exogeneity). The novel friction here is not the presence of noise per se, but that the *instrument candidate is itself part of the contract*, hence can influence selection into the observed sample. This is the sense in which "contract-as-IV" can fail even when the signal noise is classical. Put differently, the standard IV condition is not violated because the contract is correlated with the effort shock (it is not), but because conditioning on participation induces dependence between the contract and unobserved outcome shifters.

Selection, missing-not-at-random, and endogenous sampling. The fact that outcomes are observed only for participants places our model squarely in the literature on sample selection and missing data. The econometric lesson is that conditioning on sample inclusion can create bias whenever inclusion depends on unobservables that also affect outcomes. This theme runs from the classic selection model of **?** to subsequent work on selection corrections and partial identification under weaker assumptions (**?**). In our context, participation depends on the agent's private cost and outside option, and unobserved productivity enters the outcome equation. If productivity is correlated with participation-relevant components of type, then the observed sample is missing not at random: the distribution of unobserved outcome shifters among participants varies with incentives.

Our focus differs from the traditional selection-correction agenda in two ways. First, rather than proposing a parametric correction based on an explicit selection equation, we design the platform's experimental variation so that it remains orthogonal to unobservables *within the selected sample*. This is conceptually closer to designing an instrument that survives selection than to modeling selection for correction. Second, the timing of information and randomization is central. Many selection models treat the regressor variation as exogenous in the full population and then examine what happens under conditioning. Here, the platform chooses the contract, the agent chooses participation, and only then is the experimental variation realized. This sequencing allows the platform to create exogenous variation *after* selection, thereby avoiding the collider channel that would otherwise correlate the instrument with unobservables once we condition on observing outcomes.

This timing-based perspective also complements recent discussions in empirical IO and labor economics on endogenous sampling in platform data. When the platform only sees transactions that occur, and transactions oc-

cur only when agents and consumers select in, naive causal interpretation of observed variation can be badly misleading. Our framework provides a clean abstraction of this concern, isolating a minimal assumption—post-selection randomization independent of private types—under which standard moment restrictions become valid on the observed sample.

Bandits, censored feedback, and learning under participation constraints. A fourth related literature studies online learning when feedback is partial, censored, or filtered through actions. In multi-armed bandits, the learner observes outcomes only for chosen arms; in more complex partial-monitoring models, the learner observes only a signal correlated with payoffs. A growing body of work considers *censored* or *self-selecting* feedback, where rewards are observed only if an acceptance event occurs, as in dynamic pricing with demand censoring or procurement with bidder participation. These problems emphasize that exploration is constrained by the data-generating process: to learn, one must induce observations, but the act of inducing observations changes who appears and what is observed.

Our environment shares the "learning from those who show up" constraint, but with a distinctive economic structure: participation is controlled by agents, not by the learner, and the learner's action (the contract) simultaneously affects incentives and selection. This creates a two-layer endogeneity absent from standard bandits: the platform does not directly choose which data points to observe, and the distribution of observed types depends on the policy. Our results can be interpreted as providing a form of *safe exploration* tailored to this structure: by shifting randomization to occur after participation, the platform ensures that exploration happens inside the observed sample without altering the selection margin at the realization level. In bandit terms, the exploration shock affects the reward-relevant behavior conditional on being "pulled" into the sample, but does not affect the event that a sample is observed.

There is also a conceptual connection to work on bandits with constraints and strategic responses, where a learner must maintain feasibility (e.g., participation or individual rationality) while learning. In our formulation, the baseline contract plays the role of a policy lever that governs participation, while the perturbation plays the role of an exploration device whose distribution is committed to ex ante but realized after selection. The separation of these roles is precisely what prevents exploration from contaminating selection.

Platform experimentation, randomized audits, and mechanism design practice. Finally, our design is motivated by how platforms run experiments when they cannot fully randomize at the participation stage. In many settings, platforms can commit to a payment rule or scoring rule that is

publicly described, but can also introduce randomized components that are only revealed after an agent opts in: surprise bonuses, randomized weightings of score components, lottery-like incentives, or audit-based adjustments. Such features are often discussed in operational terms (fairness in expectation, budget control, robustness to gaming), but they also have a statistical role: they can create quasi-experimental variation in incentives among participants without changing who participates in response to the realized shock.

This interpretation connects our perturbation to audit regimes in mechanism design and regulation, where random audits create incentives for compliance while preserving tractability and limiting manipulation. Randomized audits are typically justified as a deterrence device; here, the same logic yields identification, because audit randomness is exogenous and can be arranged to be realized conditional on participation. Relatedly, there is a practical tension between transparency and experimentation: platforms may need to disclose the distribution of incentive schemes while keeping realizations unpredictable to prevent gaming. Our framework makes clear why such commitment matters for participation decisions (agents evaluate expected pay) and why unpredictability at the realization stage can be valuable for learning.

Summary of our contribution relative to prior work. Across these literatures, two facts are well understood: (i) selection can invalidate naive inference, and (ii) exogenous randomization can restore identification. The contribution here is to highlight a specific and implementable way to reconcile these facts in repeated contracting environments with endogenous participation. By separating the contract component that determines entry from the randomized component that generates identifying variation, and by placing the random draw after participation, we obtain a selection-robust instrument while remaining within the linear-contract paradigm commonly used in practice. This positions the subsequent model section: we formalize the timing, information, and payoff primitives under which the proposed randomization delivers valid conditional moments on the participating sample and can be embedded in an online learning-and-contracting policy.

3 Model: contracting with endogenous participation and post-entry randomization

We study a repeated contracting environment over rounds $t = 1, \dots, T$ in which a platform (the principal) interacts with a stream of short-lived agents. Each round features a fresh agent whose private type is

$$\tau_t = (c_t, u_t^0, \omega_t).$$

Here $c_t : \mathbb{R}_+^d \to \mathbb{R}_+$ is the agent's cost of effort, $u_t^0 \in \mathbb{R}$ is an outside-option utility, and $\omega_t \in \mathbb{R}$ is an agent-specific outcome shifter (interpretable as latent quality, match value, or composition). The key econometric friction is that ω_t affects the platform's realized outcome but is not observed, and may be correlated with participation-relevant components of type.

Effort, telemetry, and outcomes. If the agent participates, it chooses an effort (or action) vector $a_t \in \mathbb{R}_+^d$ across d tasks. Effort is not directly observed by the platform; instead the platform observes a noisy proxy $x_t \in \mathbb{R}^d$ (telemetry, measured behaviors, performance metrics) of the form

$$x_t \;=\; a_t + \varepsilon_t, \qquad \mathbb{E}[\varepsilon_t \mid a_t] = 0, \tag{1}$$

where ε_t is a mean-zero measurement error (typically taken to be subgaussian to support concentration later). The platform's economic outcome (revenue, value created, or some downstream KPI) is

$$y_t \;=\; \langle \theta^\star, a_t \rangle + \omega_t + \eta_t, \qquad \mathbb{E}[\eta_t \mid a_t, \omega_t] = 0, \tag{2}$$

where $\theta^\star \in \mathbb{R}_+^d$ is an unknown task-value vector that we wish to learn, and η_t is an outcome noise term (also typically subgaussian). The maintained structure in (1)–(2) is deliberately minimal: effort enters the outcome linearly via $\theta^\star$, the platform observes only a noisy proxy for effort, and an unobserved shifter ω_t affects outcomes but is not recorded.

Linear contracts with baseline slopes and randomized perturbations. In each round t, before the agent decides whether to participate, the platform posts a linear contract based on the observed telemetry x_t. The contract is parameterized by a *slope vector* in $\mathbb{R}^d$: if the agent participates, it is paid $\langle \beta_t, x_t \rangle$ for some realized slope β_t. The platform controls β_t through two components:

1. a *baseline* slope vector $b_t \in \mathcal{B} \subseteq \mathbb{R}_+^d$, where $\mathcal{B}$ is a feasible set (e.g., convex and compact, reflecting business rules, fairness constraints, or budget/market-compatibility limits); and

2. a *random perturbation* $Z_t \in \mathbb{R}^d$ with a distribution committed to ex ante, satisfying $\mathbb{E}[Z_t] = 0$ and $\mathbb{E}[Z_t Z_t^\top] = \Sigma_Z \succ 0$.

The realized contract slope is then

$$\beta_t \;=\; b_t + Z_t.$$

The role of b_t is economic and operational: it is the predictable part of incentives that agents can plan against and that the platform uses to target profitability. The role of Z_t is statistical: it injects exogenous variation

in incentive slopes among participants to support identification of $\theta^\star$ from selected-sample data. The covariance condition $\Sigma_Z \succ 0$ ensures that the perturbation spans all directions in the d-dimensional task space, ruling out degenerate experimentation that would leave some components weakly identified.

Timing and information: participation before randomization. The sequencing of moves is central. Each round proceeds as follows.

1. The platform observes its past history (formalized below as a filtration $\mathcal{F}_t$) and chooses a baseline slope $b_t \in \mathcal{B}$. It also commits to the distribution of the perturbation Z_t (we take this distribution as fixed across rounds, with known mean zero and covariance Σ_Z).

2. The agent observes b_t and the distribution of Z_t (but not its realization) and chooses whether to participate, $p_t \in \{0, 1\}$.

3. If $p_t = 1$, then Z_t is drawn and revealed, and the realized slope becomes $\beta_t = b_t + Z_t$.

4. The agent chooses effort a_t after seeing β_t.

5. Signals x_t and outcome y_t realize according to (1)–(2). The platform observes (x_t, y_t) only if $p_t = 1$.

Two observability conventions will be useful later. First, we allow the platform to observe p_t always (it sees whether an agent shows up). Second, we allow the platform to observe β_t whenever $p_t = 1$, and in our baseline formulation we also allow it to observe the realized slope even when $p_t = 0$ (since the platform can always record its own random draw); what matters for identification, however, is that (x_t, y_t) are observed only on participating rounds. The platform never observes the type τ_t nor the shifter ω_t.

Agent payoff, effort choice, and best response. If the agent participates, its utility is linear in the contract payment and quasilinear in cost:

$$U_t^A \;=\; \langle \beta_t, x_t \rangle - c_t(a_t) \;=\; \langle \beta_t, a_t \rangle + \langle \beta_t, \varepsilon_t \rangle - c_t(a_t).$$

Because ε_t has mean zero conditional on a_t, the agent's effort problem is equivalent (in expectation) to choosing a_t to maximize $\langle \beta_t, a \rangle - c_t(a)$. We assume $c_t(\cdot)$ is strictly convex and differentiable on $(\mathbb{R}_+)^d$, which delivers a unique interior best response when the optimum lies in the interior:

$$a_t \;\in\; \arg\max_{a \in \mathbb{R}_+^d} \; \langle \beta_t, a \rangle - c_t(a), \qquad \nabla c_t(a_t) = \beta_t. \tag{3}$$

It is useful to view (3) as defining an effort mapping $a(\beta_t; c_t)$: higher realized incentives β_t tilt effort toward the dimensions with larger slope, and strict

convexity prevents corner solutions from being knife-edge. We do not impose separability across tasks; allowing general convex c_t accommodates cross-task substitution and complementarity.

If the agent does not participate, it receives the outside option utility u_t^0. Thus, participation is an individual rationality decision made *before* the perturbation is realized. Writing the agent's (ex ante) participation value as the expected optimized utility under the known distribution of Z_t, we can represent participation as

$$p_t \;=\; \mathbf{1}\Big\{\mathbb{E}_Z\big[\max_{a\in\mathbb{R}_+^d}\langle b_t + Z, a\rangle - c_t(a)\big] \;\geq\; u_t^0\Big\}. \tag{4}$$

This formulation captures a salient feature of platform settings: agents decide whether to log on based on the *posted* compensation rule and their expectations about earnings, rather than on the realized shocks that occur only after they have entered.

Platform payoff and the learning problem. When the agent participates, the platform receives outcome y_t but pays $\langle \beta_t, x_t\rangle$ according to the contract. We write the platform's per-round payoff as

$$U_t^P \;=\; p_t\big(y_t - \langle \beta_t, x_t\rangle\big) \;-\; \kappa\|Z_t\|_2^2, \tag{5}$$

where $\kappa \geq 0$ is an (optional) experimentation cost that penalizes large perturbations.[1] The first term in (5) emphasizes the central tradeoff: increasing β_t can increase effort and thus y_t, but it also increases the payment through x_t; moreover, changing the baseline b_t affects participation (4), and therefore the set of observed outcomes.

The platform does not know $\theta^\star$ a priori, so it cannot directly choose b_t to maximize long-run profit. Instead it must learn $\theta^\star$ from the data it observes. Crucially, data are available only on the set of participating rounds

$$\mathcal{T}_1 \;=\; \{t \in \{1,\ldots,T\} : p_t = 1\}.$$

This censoring is endogenous: which rounds enter $\mathcal{T}_1$ depends on the posted baseline b_t and on the distribution of Z_t through the agent's expected utility, as well as on the unobserved type sequence.

Filtration and what is observable to the platform. Let $\mathcal{F}_t$ denote the platform's information before choosing b_t in round t. At a minimum, $\mathcal{F}_t$ contains past choices and realized data on participating rounds:

$$\mathcal{F}_t \;=\; \sigma\Big(\{(b_s, p_s, \beta_s, \mathbf{1}\{p_s = 1\}x_s, \mathbf{1}\{p_s = 1\}y_s)\}_{s<t}\Big).$$

[1] One interpretation is that volatile incentives are costly: they may harm agent trust, increase churn, or create fairness concerns. Setting $\kappa = 0$ isolates the statistical role of Z_t.

This makes explicit that the platform learns from a history in which some covariates are systematically missing when $p_s = 0$. Our policies will be $\mathcal{F}_t$-measurable mappings into $\mathcal{B}$: the platform chooses b_t based on past observed outcomes and signals, anticipating how b_t influences future participation and effort.

Benchmark and regret. To evaluate an online policy, we compare its cumulative payoff to that of the best fixed baseline slope in hindsight. Concretely, fix the perturbation distribution (hence Σ_Z) and consider any $b \in \mathcal{B}$ used as a constant baseline across time, with the same timing, participation rule (4), effort best response (3), and payoff (5). We define the (pseudo-)regret of a policy $\pi = \{b_t\}_{t=1}^T$ as

$$\mathrm{Regret}(T) \;=\; \max_{b\in\mathcal{B}} \; \mathbb{E}\Big[\sum_{t=1}^{T} U_t^P(b)\Big] \;-\; \mathbb{E}\Big[\sum_{t=1}^{T} U_t^P(b_t)\Big],$$

where the expectation is taken over the perturbations and noise terms (and, when appropriate, over any randomness in the policy). This notion isolates the learning component: even if the optimal baseline is not implementable without knowing $\theta^\star$, a good policy should approach its performance as data accumulate. At the same time, the regret criterion respects the participation constraint embedded in the environment, because payoffs and observations are generated only when agents choose $p_t = 1$ under the posted baseline and the known perturbation distribution.

The object of interest in the next sections is how to learn $\theta^\star$ and optimize b_t in this censored-feedback environment. The central challenge is that, while linear contracts naturally suggest using contract variation to identify the effect of effort on outcomes, the baseline contract also shifts participation and hence the composition of observed types. The post-participation perturbation Z_t is designed to disentangle these forces by creating within-participant variation that is orthogonal to the unobserved shifter ω_t even after conditioning on $p_t = 1$.

4 Failure of contract-as-IV under selection

A natural first impulse in linear contracting models is to treat contract variation as an instrument for effort. In our setting the platform observes (x_t, y_t) only when $p_t = 1$, and it is tempting to estimate $\theta^\star$ by IV on the participating sample using either the posted baseline b_t or the realized slope β_t as an instrument for the endogenous regressor x_t. Formally, one might hope that $\theta^\star$ satisfies a moment restriction of the form

$$\mathbb{E}\big[q_t\big(y_t - \langle\theta^\star, x_t\rangle\big) \mid p_t = 1\big] = 0, \tag{6}$$

where $q_t \in \{b_t, \beta_t\}$ (or some function thereof). This section explains why (6) typically fails under endogenous participation, even though (i) the measurement error ε_t is mean-zero and (ii) the platform sets incentives before observing any round-t shocks. The core issue is that b_t shifts who enters, and the unobserved outcome shifter ω_t can move with entry in ways that correlate with the instrument in the selected sample.

4.1 A two-type example: composition shifts generate spurious "first-stage–residual" correlation

We illustrate the failure most transparently in a one-dimensional specialization ($d = 1$) with no perturbation. Let $Z_t \equiv 0$, so $\beta_t = b_t \equiv b$. Suppose effort is chosen after observing $\beta = b$ and costs are quadratic:

$$c(a) = \frac{1}{2}a^2, \qquad a \in \mathbb{R}_+.$$

Then the agent's best response is interior and given by $a = b$, yielding optimized (expected) participation value

$$\max_{a \geq 0}\{ba - \tfrac{1}{2}a^2\} = \frac{1}{2}b^2.$$

Consider two types indexed by $k \in \{L, H\}$, arriving i.i.d. across rounds with $\Pr(k = H) = \pi \in (0, 1)$. Types differ in both outside option and latent quality:

$$u_L^0 = \underline{u}, \quad u_H^0 = \overline{u}, \quad \overline{u} > \underline{u}, \qquad \omega_L = 0, \quad \omega_H = \Delta > 0.$$

Thus the high-quality type also has a higher outside option. Participation is determined before any outcome noise is realized, and in this simple case it is deterministic given b:

$$p = \mathbf{1}\left\{\tfrac{1}{2}b^2 \geq u_k^0\right\}.$$

Let $\underline{b} = \sqrt{2\underline{u}}$ and $\overline{b} = \sqrt{2\overline{u}}$, so that for $b \in [\underline{b}, \overline{b})$ only low types participate, while for $b \geq \overline{b}$ both types participate.

Now consider the platform's outcome and telemetry:

$$x = a + \varepsilon = b + \varepsilon, \qquad y = \theta^\star a + \omega + \eta = \theta^\star b + \omega + \eta,$$

with $\mathbb{E}[\varepsilon \mid a] = 0$ and $\mathbb{E}[\eta \mid a, \omega] = 0$. The structural residual at the truth is

$$y - \theta^\star x = \omega + \eta - \theta^\star \varepsilon.$$

Unconditionally (i.e., absent selection) the moment $\mathbb{E}[b(y - \theta^\star x)]$ would be driven by $\mathbb{E}[b\omega]$, and could be zero if b were independent of ω and $\mathbb{E}[\omega] = 0$ or if ω were mean-zero noise. But we do not observe unconditionally; we

observe only when $p = 1$. Conditioning on $p = 1$ changes the distribution of ω as a function of b:

$$\mathbb{E}[\omega \mid p = 1, b] = \begin{cases} 0, & b \in [\underline{b}, \bar{b}), \\ \pi\Delta, & b \geq \bar{b}, \end{cases}$$

because the high type appears in the participating sample only when b is high enough to clear its outside option. Consequently, on the participating sample the naive IV moment with instrument b satisfies

$$\begin{aligned} \mathbb{E}\big[b\big(y - \theta^\star x\big) \mid p = 1, b\big] &= b\,\mathbb{E}[\omega \mid p = 1, b] \;+\; b\,\mathbb{E}[\eta \mid p = 1, b] \;-\; \theta^\star b\,\mathbb{E}[\varepsilon \mid p = 1, b] \\ &= b\,\mathbb{E}[\omega \mid p = 1, b], \end{aligned} \tag{7}$$

where the last equality uses mean-zero assumptions for η and ε together with the fact that participation is decided before η and ε are realized. By the expression above, (7) is strictly positive whenever $b \geq \bar{b}$:

$$\mathbb{E}\big[b\big(y - \theta^\star x\big) \mid p = 1, b\big] = b\,\pi\Delta > 0.$$

Thus $\theta^\star$ does *not* solve the selected-sample moment restriction (6) when $q = b$. Econometrically, the instrument is correlated with the residual through a composition channel: raising b mechanically improves the latent quality mix among participants, so the outcome residual increases with b even when the structural relationship between a and y is correctly specified.

The same logic shows why using the realized contract slope β as the instrument is not a remedy when β contains a baseline component that shifts participation. In the current example $\beta = b$, so nothing changes. More generally, even when we later add mean-zero perturbations, the realized slope $\beta_t = b_t + Z_t$ inherits the selection-induced correlation coming from the predictable component b_t.

4.2 General bias decomposition: where the naive moment fails

We now isolate the general mechanism in the full d-dimensional model. Fix a candidate $\theta \in \mathbb{R}^d$ and define the regression residual

$$r_t(\theta) \;=\; y_t - \langle \theta, x_t \rangle. \tag{8}$$

Substituting $x_t = a_t + \varepsilon_t$ and $y_t = \langle \theta^\star, a_t \rangle + \omega_t + \eta_t$ yields

$$r_t(\theta) = \langle \theta^\star - \theta, a_t \rangle + \omega_t + \eta_t - \langle \theta, \varepsilon_t \rangle. \tag{9}$$

At the truth $\theta = \theta^\star$,

$$r_t(\theta^\star) = \omega_t + \eta_t - \langle \theta^\star, \varepsilon_t \rangle. \tag{10}$$

Consider an instrument q_t that is measurable with respect to the platform's information at the contracting stage (e.g., $q_t = b_t$ or $q_t = \beta_t$). On the selected sample we would like

$$\mathbb{E}[q_t\, r_t(\theta^\star) \mid p_t = 1] = 0. \tag{11}$$

Using (10), we can decompose the left-hand side as

$$\mathbb{E}[q_t\, \omega_t \mid p_t = 1] \;+\; \mathbb{E}[q_t\, \eta_t \mid p_t = 1] \;-\; \mathbb{E}[q_t\, \langle \theta^\star, \varepsilon_t \rangle \mid p_t = 1]\,. \tag{12}$$

Under our maintained assumptions, the second and third terms are benign: η_t and ε_t are mean-zero conditional on primitives determined before their realization, and participation is chosen prior to these noises. In particular, for any q_t measurable before η_t is drawn, $\mathbb{E}[q_t\, \eta_t \mid p_t = 1] = 0$; similarly, $\mathbb{E}[q_t\, \varepsilon_t \mid p_t = 1] = 0$ when $\mathbb{E}[\varepsilon_t \mid a_t] = 0$ and q_t is measurable prior to ε_t.[2]

The problematic term is therefore

$$\mathbb{E}[q_t\, \omega_t \mid p_t = 1]\,. \tag{13}$$

If we had full participation ($p_t \equiv 1$), then (13) would reduce to $\mathbb{E}[q_t \omega_t]$, which can be made zero under standard exogeneity assumptions (e.g., q_t randomized independently of ω_t). Under endogenous participation, however, we are conditioning on an event whose probability depends on both q_t (through incentives) and the agent's type (through costs and outside options), and ω_t may be statistically linked to those participation-relevant type components. Applying iterated expectations makes the channel explicit:

$$\mathbb{E}[q_t\, \omega_t \mid p_t = 1] = \mathbb{E}[q_t \cdot \mathbb{E}[\omega_t \mid p_t = 1, q_t] \mid p_t = 1]\,. \tag{14}$$

Unless $\mathbb{E}[\omega_t \mid p_t = 1, q_t]$ is constant in q_t, the inner conditional mean varies with the instrument, and the product in (14) generally does not average to zero. The two-type example above produces exactly such variation: higher incentives expand participation to include higher-ω agents.

This highlights the precise sense in which "contract-as-IV" fails. Even if the instrument is set without observing ω_t, the instrument changes the *composition* of ω_t in the observed data because selection depends on the contract. Put differently, the exclusion restriction must hold *after conditioning on being observed*, and the conditioning event is itself a function of the contract.

4.3 When do the source moment conditions hold, and when do they fail?

The moment conditions used in standard linear IV arguments implicitly presume that the instrument is orthogonal to the structural residual in the

[2] One can make this argument formal by conditioning on (a_t, p_t, q_t) and applying iterated expectations. The key is that the selection event $p_t = 1$ is decided before ε_t and η_t are realized, so selection does not induce nonzero means in these noise terms.

estimation sample. In our environment the estimation sample is endogenous, so orthogonality must be assessed conditional on $p_t = 1$. The naive moment (6) can hold in special cases, each corresponding to the failure of at least one link in the selection channel:

1. *No selection / full participation:* if $p_t \equiv 1$ (e.g., outside options are always low enough), then the selected-sample and population moments coincide.

2. *No latent shifter:* if $\omega_t \equiv 0$, then selection affects observability but does not create omitted-variable correlation with the instrument.

3. *Latent shifter independent of participation-relevant type components:* if ω_t is independent of (c_t, u_t^0) (or more generally independent of the agent's participation decision given the posted contract), then the conditional mean $\mathbb{E}[\omega_t \mid p_t = 1, q_t]$ does not vary with q_t.

4. *Instrument does not affect entry:* if q_t shifts incentives within the participant pool but does not change the participation event, then conditioning on $p_t = 1$ does not induce correlation between q_t and ω_t. This is exactly the logic we exploit later by drawing certain randomizations *after* participation.

Outside these knife-edge or design-driven cases, however, the selection term (13) is present. In particular, using b_t as an instrument is typically invalid precisely because b_t is posted before participation and is the primary driver of entry. Using β_t as an instrument does not generally fix the issue either: β_t contains b_t, and hence inherits the component that moves composition. Formally, writing $\beta_t = b_t + Z_t$ and taking $q_t = \beta_t$ in (13) yields

$$\mathbb{E}[\beta_t \omega_t \mid p_t = 1] = \mathbb{E}[b_t \omega_t \mid p_t = 1] + \mathbb{E}[Z_t \omega_t \mid p_t = 1],$$

and even if the perturbation part is orthogonal, the baseline part need not be.

Finally, it is worth emphasizing a distinct failure mode that motivates our timing restriction. If randomization were realized *before* the participation decision (so that p_t depended on the realized slope rather than only on its distribution), then even a mean-zero perturbation could become correlated with ω_t in the selected sample, since the selection event would depend directly on the realized random draw. The lesson is that exogeneity of the instrument must be evaluated *conditional on the realized sample*, and the timing of randomization determines whether conditioning on $p_t = 1$ induces correlation.

Taken together, these observations motivate the need for instruments that generate within-sample variation in incentives while being insulated from the participation margin. In the next section we describe designs that satisfy this requirement and deliver valid conditional moments despite endogenous entry.

5 Safe instruments under endogenous participation

Our diagnosis in Section 4 points to a design principle: the instrument must generate variation in incentives *within* the observed (participating) sample while remaining insulated from the participation margin. Because participation is itself a function of the posted contract, any object realized prior to entry is a potential driver of composition and hence can inherit correlation with the unobserved outcome shifter ω_t in the selected sample. In contrast, randomizations realized *after* the agent commits to participate do not affect who enters, and therefore can remain orthogonal to latent quality even though the observed sample is selected.

We describe two concrete families of such "safe" instruments. The first is the post-participation slope perturbation Z_t already built into our baseline model. The second implements the same logic using randomized audits or repeated measurements, which can be useful when operational constraints limit how we can randomize the primary contract slope on the main telemetry.

5.1 Post-participation slope randomization

The platform posts a baseline slope $b_t \in \mathcal{B}$ and commits to a distribution for a mean-zero perturbation Z_t with covariance $\Sigma_Z \succ 0$. The timing restriction is that the agent chooses participation based on $(b_t, \mathcal{L}(Z_t))$, while the realization of Z_t is drawn and revealed only *after* p_t is chosen. If $p_t = 1$, the realized linear payment slope is $\beta_t = b_t + Z_t$, effort is chosen as

$$a_t \in \arg\max_{a \in \mathbb{R}^d_+} \{\langle \beta_t, a\rangle - c_t(a)\},$$

and the platform observes (x_t, y_t) with $x_t = a_t + \varepsilon_t$ and $y_t = \langle \theta^\star, a_t\rangle + \omega_t + \eta_t$.

The key property is that, conditional on $p_t = 1$, the perturbation remains exogenous with respect to unobservables that enter the outcome equation.

Exogeneity condition. We require that, for each t, the random vector Z_t is independent of $(\tau_t, \varepsilon_t, \eta_t)$ and hence of $(c_t, u_t^0, \omega_t, \varepsilon_t, \eta_t)$, and that $\mathbb{E}[Z_t] = 0$. Because p_t is measurable with respect to $(b_t, \mathcal{L}(Z_t), \tau_t)$ and is chosen before Z_t is realized, the selection event $\{p_t = 1\}$ depends on τ_t and the *distribution* of Z_t but not on its realization. As a result, conditioning on $p_t = 1$ does not create dependence between Z_t and ω_t (or any other shock), and mean-zero is preserved:

$$\mathbb{E}[Z_t \mid p_t = 1] = \mathbb{E}[Z_t] = 0, \qquad \mathbb{E}[Z_t\, \omega_t \mid p_t = 1] = \mathbb{E}[Z_t] \cdot \mathbb{E}[\omega_t \mid p_t = 1] = 0.$$

To translate this into an estimable moment, define the residual at a candidate parameter θ by $r_t(\theta) = y_t - \langle \theta, x_t\rangle$. At the truth, $r_t(\theta^\star) = \omega_t +$

$\eta_t - \langle \theta^\star, \varepsilon_t \rangle$, so for participating rounds,

$$\begin{aligned}\mathbb{E}[Z_t\, r_t(\theta^\star) \mid p_t = 1] &= \mathbb{E}[Z_t\, \omega_t \mid p_t = 1] + \mathbb{E}[Z_t\, \eta_t \mid p_t = 1] - \mathbb{E}[Z_t\, \langle \theta^\star, \varepsilon_t \rangle \mid p_t = 1] \\ &= 0, \end{aligned} \tag{15}$$

where the last line uses (i) $Z_t \perp \omega_t$ even after conditioning on $p_t = 1$, (ii) $Z_t \perp \eta_t$ and $\mathbb{E}[\eta_t \mid a_t, \omega_t] = 0$, and (iii) $Z_t \perp \varepsilon_t$ and $\mathbb{E}[\varepsilon_t \mid a_t] = 0$.

Relevance condition. Exogeneity alone is not enough; we also need Z_t to move x_t in the selected sample. The corresponding rank condition is that the cross-moment matrix

$$M \;\equiv\; \mathbb{E}[Z_t x_t^\top \mid p_t = 1]$$

is nonsingular. Since $x_t = a_t + \varepsilon_t$ and ε_t is independent of Z_t with mean zero, we have

$$\mathbb{E}[Z_t x_t^\top \mid p_t = 1] = \mathbb{E}[Z_t a_t^\top \mid p_t = 1]. \tag{16}$$

Thus relevance is governed by how the agent's best response $a(\beta; c)$ co-moves with the realized perturbation Z_t. In many canonical cases relevance is immediate: for example, with separable quadratic costs $c_t(a) = \frac{1}{2}\|a\|_2^2$, we obtain $a_t = \beta_t = b_t + Z_t$, and hence $M = \mathbb{E}[Z_t(b_t + Z_t)^\top \mid p_t = 1] = \Sigma_Z$, which is invertible by assumption.

More generally, strict convexity and differentiability of c_t imply that the (interior) best response satisfies $\nabla c_t(a_t) = \beta_t$, so a_t is a (typically monotone) transformation of $\beta_t = b_t + Z_t$. If this mapping is sufficiently responsive in all directions and the support of Z_t spans $\mathbb{R}^d$ (captured by $\Sigma_Z \succ 0$), then the matrix in (16) is generically full rank on the participating sample. The substantive failure mode is weak instruments: if Σ_Z is nearly singular, or if costs are such that actions are insensitive to some coordinates of β_t, then some components of $\theta^\star$ will be weakly identified.

5.2 Randomized audits and repeated measurements as instrument generators

In some applications it is operationally difficult to add random perturbations directly to the main performance slope (e.g., due to fairness constraints, regulatory limits on pay variance, or product constraints on the displayed formula). A useful alternative is to create exogenous, post-participation variation in *marginal incentives* via an auxiliary measurement or audit channel whose realization is randomized after participation. The underlying econometric logic is the same: the instrument must (i) be realized after entry so it does not shift composition, and (ii) affect the agent's chosen effort so it is relevant for x_t.

A generic audit design. Suppose that if $p_t = 1$ the platform can (with some probability) conduct an audit that produces an auxiliary signal $\tilde{x}_t \in \mathbb{R}^d$ satisfying

$$\tilde{x}_t = a_t + \tilde{\varepsilon}_t, \qquad \mathbb{E}[\tilde{\varepsilon}_t \mid a_t] = 0,$$

with $\tilde{\varepsilon}_t$ independent of (ε_t, η_t) and of the agent type. After participation, the platform draws a randomized *audit bonus slope* W_t with $\mathbb{E}[W_t] = 0$ and announces that the total payment will be

$$\langle b_t, x_t \rangle + \langle W_t, \tilde{x}_t \rangle. \tag{17}$$

The agent then chooses effort after observing W_t (together with the realized audit signal rule), so the realized marginal incentives become $b_t + W_t$ on the audited channel. Importantly, the random variable W_t is drawn after participation, so it does not affect p_t and does not induce composition shifts.

Under the analogue of the independence and mean-zero conditions imposed on Z_t, we obtain the same selected-sample orthogonality:

$$\mathbb{E}\big[W_t\big(y_t - \langle \theta^\star, x_t \rangle\big) \mid p_t = 1\big] = 0.$$

Relevance is again a rank condition, now involving $\mathbb{E}[W_t x_t^\top \mid p_t = 1]$. The practical advantage of (17) is that, when the main slope b_t must remain stable, one can still inject experimentally useful variation through the audit bonus while keeping expected payments (and hence entry incentives) unchanged in expectation.

Repeated measurements. A closely related variant replaces "audit" with repeated measurement: conditional on participation the platform obtains $K \geq 2$ independent proxies

$$x_t^{(k)} = a_t + \varepsilon_t^{(k)}, \qquad k = 1, \ldots, K,$$

and randomizes (after participation) a weight vector $(\lambda_t^{(1)}, \ldots, \lambda_t^{(K)})$ with mean $(1, 0, \ldots, 0)$ (or any fixed baseline allocation) and $\sum_k \lambda_t^{(k)} = 1$. The realized payment depends on $\sum_k \lambda_t^{(k)} x_t^{(k)}$, which the agent observes before choosing effort. The deviation $\lambda_t^{(k)} - \mathbb{E}[\lambda_t^{(k)}]$ plays the role of a mean-zero instrument that shifts the marginal incentive placed on each measurement realization without changing the expected contract. This design can be attractive when each individual measurement channel is noisy or manipulable, yet the platform can vary which channel is "paid on" without altering entry decisions.

Across these implementations the common requirement is that the randomization is (i) realized after participation and (ii) independent of latent quality and shocks, so that conditioning on $p_t = 1$ does not contaminate the exclusion restriction.

5.3 Identification from selected-sample moments

We summarize the preceding discussion as an identification result stated directly on the selected sample, since (x_t, y_t) are observed only when $p_t = 1$. Let W_t denote any post-participation randomized instrument (e.g., $W_t = Z_t$ in the baseline model, or an audit-bonus randomization as above) such that $\mathbb{E}[W_t] = 0$ and W_t is independent of $(\tau_t, \varepsilon_t, \eta_t)$, with the additional timing requirement that p_t is chosen before W_t is realized.

Theorem (Selected-sample IV identification). Suppose (i) W_t is drawn after participation and is independent of $(\tau_t, \varepsilon_t, \eta_t)$ with $\mathbb{E}[W_t] = 0$, and (ii) the relevance matrix $\mathbb{E}[W_t x_t^\top \mid p_t = 1]$ is nonsingular. Then $\theta^\star$ is the unique solution to the conditional moment restriction

$$\mathbb{E}\big[W_t\big(y_t - \langle \theta, x_t \rangle\big) \mid p_t = 1\big] = 0. \tag{18}$$

Discussion. To see uniqueness, expand (18) as

$$\mathbb{E}[W_t y_t \mid p_t = 1] - \mathbb{E}[W_t x_t^\top \mid p_t = 1]\theta = 0,$$

so nonsingularity yields the closed-form solution

$$\theta = \Big(\mathbb{E}[W_t x_t^\top \mid p_t = 1]\Big)^{-1} \mathbb{E}[W_t y_t \mid p_t = 1],$$

and exogeneity ensures that $\theta = \theta^\star$ satisfies the equation. The substantive content is that the orthogonality is asserted *after* conditioning on participation; this is exactly where baseline contracts fail as instruments and where post-participation randomization succeeds.

Finally, we emphasize the tradeoff implicit in designing W_t: stronger randomization improves relevance (and hence statistical precision) but can reduce contemporaneous surplus by distorting incentives away from the baseline optimum, and may depress participation if agents are sufficiently risk averse or face downside risk in the realized slope. These considerations become central once we turn to estimation and online policy design, where the platform must choose the magnitude of randomization to balance learning speed against short-run welfare and participation.

5.4 Estimation on selected samples: IV/GMM, finite-sample guarantees, and design diagnostics

Having established the selected-sample moment condition (18), we now turn to estimation using only the rounds in which the agent participates. This step is not merely a technicality: selection affects the *effective* sample size, and (through action responsiveness) it also affects the *strength* of the instrument, both of which enter directly into finite-sample performance.

Sample moments and the IV/GMM estimator. Let $\mathcal{T}_1 \equiv \{t \le T : p_t = 1\}$ denote the set of participating rounds and $n \equiv |\mathcal{T}_1|$ its cardinality. For any post-participation instrument W_t satisfying the conditions of the preceding subsection, the identification moment is

$$\mathbb{E}\big[W_t\big(y_t - \langle \theta, x_t\rangle\big) \mid p_t = 1\big] = 0.$$

A natural estimator replaces this conditional expectation by its empirical analogue on $\mathcal{T}_1$:

$$\frac{1}{n}\sum_{t\in\mathcal{T}_1} W_t\big(y_t - \langle \theta, x_t\rangle\big) = 0. \tag{19}$$

When $W_t \in \mathbb{R}^d$ (as in slope perturbations or audit-bonus vectors), (19) yields d equations in d unknowns. Writing $X \in \mathbb{R}^{n\times d}$ for the matrix with rows $x_t^\top$, $W \in \mathbb{R}^{n\times d}$ for the matrix with rows $W_t^\top$, and $Y \in \mathbb{R}^n$ for the vector with entries y_t, the just-identified IV estimator is the familiar closed form

$$\hat\theta = \big(W^\top X\big)^{-1} W^\top Y, \tag{20}$$

provided $W^\top X$ is nonsingular.[3]

In settings with more instruments than regressors (e.g., if we stack multiple audit realizations, multiple perturbations, or interaction terms), we can adopt standard GMM on the selected sample. Let $m_t(\theta) \equiv W_t\big(y_t - \langle \theta, x_t\rangle\big) \in \mathbb{R}^q$ with $q \ge d$. The selected-sample GMM estimator solves

$$\hat\theta_{\mathrm{GMM}} \in \arg\min_{\theta\in\mathbb{R}^d} \left\| \frac{1}{n}\sum_{t\in\mathcal{T}_1} m_t(\theta) \right\|^2_{\hat\Omega^{-1}}, \qquad \|v\|^2_{\hat\Omega^{-1}} \equiv v^\top \hat\Omega^{-1} v,$$

where $\hat\Omega$ is a consistent estimate of the selected-sample covariance of $m_t(\theta^\star)$. The conceptual message is unchanged: all moments and weighting are computed *within* the participating sample, because that is where the orthogonality is guaranteed and where data exist.

Finite-sample error: what drives statistical precision under selection. Define the true residual (at $\theta^\star$) by

$$\Gamma_t \;\equiv\; y_t - \langle \theta^\star, x_t\rangle \;=\; \omega_t + \eta_t - \langle \theta^\star, \varepsilon_t\rangle, \qquad (p_t = 1).$$

For the just-identified estimator (20), a single algebraic rearrangement yields

$$\hat\theta - \theta^\star = \big(W^\top X\big)^{-1} W^\top \Gamma, \tag{21}$$

where $\Gamma \in \mathbb{R}^n$ stacks Γ_t over $t \in \mathcal{T}_1$. Equation (21) isolates two objects that determine finite-sample performance:

[3] Equivalently, we may write sums with the indicator p_t as $\sum_{t\le T} p_t W_t x_t^\top$ and $\sum_{t\le T} p_t W_t y_t$, which makes explicit that nonparticipating rounds contribute neither signals nor outcomes.

1. the *noise term* $W^\top \Gamma = \sum_{t \in \mathcal{T}_1} W_t \Gamma_t$, which is well-behaved because W_t is mean-zero and independent of the shocks entering Γ_t even after conditioning on $p_t = 1$; and

2. the *instrument strength matrix* $W^\top X = \sum_{t \in \mathcal{T}_1} W_t x_t^\top$, which can be ill-conditioned if the perturbations are too small, too collinear, or if actions are insufficiently responsive in some dimensions.

Under standard subgaussian/boundedness conditions on W_t and Γ_t (conditional on the principal's filtration and on $p_t = 1$), martingale concentration tools yield high-probability bounds of the schematic form

$$\|\hat{\theta} - \theta^\star\|_2 \;\le\; C \, \frac{\sqrt{d\, n \, \log(dn/\delta)}}{\sigma_{\min}(W^\top X)} \tag{22}$$

with probability at least $1-\delta$, for a constant C depending on tail parameters. Two implications are immediate and economically meaningful. First, selection enters only through $n = |\mathcal{T}_1|$: lower participation slows learning because it literally reduces the number of usable moments. Second, selection also enters through $\sigma_{\min}(W^\top X)$: even if participation is high, an instrument that fails to generate within-sample incentive variation (or generates variation in only a few directions) yields weak identification and large error.

In practice, we often prefer a *regularized* version of (20) to guard against near-singularity:

$$\hat{\theta}_\lambda = \big(W^\top X + \lambda I\big)^{-1} W^\top Y, \qquad \lambda > 0, \tag{23}$$

which trades small bias for stability when $W^\top X$ is ill-conditioned. This is especially useful early in the horizon, when n is small and the realized perturbations may not yet span $\mathbb{R}^d$ in a numerically robust way.

Weak-instrument diagnostics in the selected sample. Because our identification argument hinges on relevance within the participating sample, weak-instrument concerns must also be assessed within that sample. Operationally, we recommend tracking diagnostics that are direct functions of $W^\top X$ (or its normalized analogue). The simplest is the minimum singular value $\sigma_{\min}(W^\top X)$ appearing in (22); a closely related scale-free quantity is the condition number

$$\kappa(W^\top X) \equiv \frac{\sigma_{\max}(W^\top X)}{\sigma_{\min}(W^\top X)}.$$

When $\sigma_{\min}(W^\top X)$ is small or $\kappa(W^\top X)$ is large, inference can be unstable and confidence intervals can be misleading if one relies on asymptotic approximations.

Two features of our environment deserve emphasis. First, weak instruments can arise even when $\Sigma_Z \succ 0$ in the design, because relevance is mediated by behavior: if costs make some components of a_t nearly insensitive to incentives, then x_t will not load on those components of W_t. Second, weak instruments can be *endogenous to policy*: if the baseline b_t is chosen so that participants concentrate in a region where actions saturate (e.g., corner solutions due to nonnegativity constraints or technological limits), then within-sample responsiveness can collapse, again degrading $W^\top X$.

When diagnostics indicate weakness, the natural remedies mirror classical IV practice but must respect our timing constraint. We can increase the variance of post-participation perturbations, adjust their covariance to better span under-identified directions, enrich the instrument set (e.g., multiple independent perturbations), or adopt weak-IV-robust inference procedures (such as Anderson–Rubin-type tests) computed on $\mathcal{T}_1$. The key point is that all such interventions are feasible without reintroducing selection bias so long as the randomization is realized after participation.

Choosing the randomization variance: a precision–welfare tradeoff. The ability to tune the distribution of W_t (or Z_t in the baseline model) creates a design problem: more randomization strengthens identification, but it can reduce contemporaneous surplus by distorting incentives away from the baseline and may impose explicit experimentation costs (e.g., the $-\kappa\|Z_t\|_2^2$ term).

A useful way to formalize this tradeoff is to separate *statistical* and *economic* effects of scaling. Consider, for simplicity, a homoscedastic design $Z_t = \sigma\xi_t$ with $\mathbb{E}[\xi_t] = 0$ and $\mathbb{E}[\xi_t\xi_t^\top] = I$. In many smooth environments the relevance matrix scales approximately linearly in σ:

$$\mathbb{E}[Z_t x_t^\top \mid p_t = 1] \approx \sigma\, \mathbb{E}[\xi_t a_t^\top \mid p_t = 1],$$

so the "first-stage" strength $\sigma_{\min}(\mathbb{E}[Z_t x_t^\top \mid p_t = 1])$ is (locally) increasing in σ. Holding n fixed, (22) then suggests an estimation error that shrinks roughly like $1/\sigma$. On the other hand, the contemporaneous payoff loss from randomization typically grows like σ^2 under a second-order approximation: since $\mathbb{E}[Z_t] = 0$, the first-order effect of perturbing the slope cancels, while curvature generates a quadratic welfare cost. If we also include the explicit experimentation penalty, then $\mathbb{E}[\kappa\|Z_t\|_2^2] = \kappa\sigma^2\mathbb{E}\|\xi_t\|_2^2$.

This back-of-the-envelope calculus yields a familiar shape: the marginal benefit of increasing σ (in estimation precision) diminishes, while the marginal cost (in distortion and experimentation expense) increases. In finite horizons, this suggests policies that randomize more early on (when learning is valuable) and reduce randomization later (when exploitation dominates), subject to the additional constraint that randomization must not endanger participation. Although our baseline model abstracts from risk aversion, in

many applications agents face risk or downside constraints, and large realized slopes can depress entry even if $\mathbb{E}[Z_t] = 0$. Thus, beyond variance, the *support* and tail behavior of the perturbation distribution matter for feasibility.

A practical design rule is therefore to (i) choose the shape (covariance) of Z_t to target directions where responsiveness is empirically weak, (ii) cap tails to respect operational constraints and participation stability, and (iii) select a scale σ that keeps the observed instrument-strength diagnostics comfortably away from degeneracy while limiting short-run distortion. This empirical tuning sets the stage for the online learning problem: once we repeatedly update b_t based on accumulating selected-sample IV estimates, we must jointly manage learning, payoff, and participation constraints over time.

5.5 Online learning with safe post-participation randomization

We now integrate the selected-sample IV estimator into an online contract policy. The economic challenge is that the baseline slope b_t affects both (i) the contemporaneous payoff through the induced effort choice and (ii) the future quality of statistical information by changing which agents participate. The statistical challenge is that we only observe (x_t, y_t) on the endogenous subset $\mathcal{T}_1$, so learning rates must be expressed in terms of the *realized* number of participants and the realized strength of the within-sample first stage.

A learning objective and regret benchmark. Let the platform's one-round payoff be

$$U_t^P(b_t, Z_t) \;\equiv\; p_t\Big(y_t - \langle b_t + Z_t, x_t\rangle\Big) - \kappa\|Z_t\|_2^2,$$

and define the conditional expected payoff of a baseline b (given the policy's filtration $\mathcal{F}_t$) by

$$\mu_t(b) \;\equiv\; \mathbb{E}\big[U_t^P(b, Z_t) \mid \mathcal{F}_t\big]\,,$$

where the expectation is taken over the arriving agent and all contemporaneous shocks, including Z_t drawn after participation. We evaluate a policy against the best *fixed* feasible baseline in hindsight,

$$b^\star \in \arg\max_{b\in\mathcal{B}_{\text{safe}}} \sum_{t=1}^{T} \mu_t(b),$$

where $\mathcal{B}_{\text{safe}} \subseteq \mathcal{B}$ is a subset of baselines that preserve participation feasibility in the sense described below. The (pseudo-)regret is

$$\mathrm{Reg}(T) \;\equiv\; \sum_{t=1}^{T} \Big(\mu_t(b^\star) - \mu_t(b_t)\Big).$$

This benchmark is intentionally modest: it holds the baseline fixed and does not credit policies for tracking nonstationarity in types. Nonetheless it captures the core learning–exploitation tension while remaining compatible with endogenous participation and partial observability.

A concrete policy: "estimate $\theta^\star$, then plug in." We consider policies with two coupled components: (i) an estimator $\hat{\theta}_t$ built only from participating rounds, using the post-participation perturbations Z_s as instruments; and (ii) a baseline update rule $b_t = \phi_t(\hat{\theta}_t)$ that maps beliefs about $\theta^\star$ into a contract.

A simple instantiation is the following.

1. *Experiment design (fixed shape, time-varying scale).* Fix a covariance shape $\Sigma_Z \succ 0$ and draw $Z_t = \sigma_t \xi_t$ where ξ_t is mean-zero with $\mathbb{E}[\xi_t \xi_t^\top] = \Sigma_Z$ and is independent of $(\tau_t, \varepsilon_t, \eta_t)$. The scalar σ_t can be scheduled (e.g., decreasing) to manage the precision–distortion tradeoff.

2. *Selected-sample IV update.* Let $\mathcal{T}_{1,t-1} = \{s < t : p_s = 1\}$ and $n_{t-1} = |\mathcal{T}_{1,t-1}|$. Form the regularized estimator

$$\hat{\theta}_t = \left(\sum_{s \in \mathcal{T}_{1,t-1}} Z_s x_s^\top + \lambda I \right)^{-1} \left(\sum_{s \in \mathcal{T}_{1,t-1}} Z_s y_s \right), \qquad \lambda > 0. \tag{24}$$

3. *Baseline update with projection/safety.* Choose

$$b_t \;=\; \Pi_{\mathcal{B}_{\text{safe}}}\big(\phi(\hat{\theta}_t)\big), \tag{25}$$

   where $\phi(\cdot)$ is a smooth, Lipschitz map (e.g., $\phi(\theta) = \theta/k$ for a scaling constant $k > 0$) and Π denotes Euclidean projection.

The policy is deliberately modular: (24) isolates the statistical step (which is robust to selection because Z_t is realized after entry), while (25) isolates the economic step of translating task values into incentives.

Participation feasibility and "safe" baselines. Endogenous participation is not merely a nuisance: if a learning rule drives entry to zero, the platform loses both payoff and data, and the IV estimator stops updating. We therefore impose a feasibility notion that rules out such pathological trajectories.

Because the platform does not observe the agent's cost function c_t or outside option u_t^0, we cannot directly enforce the participation constraint

$$\mathbb{E}_Z \left[\max_{a \in \mathbb{R}_+^d} \langle b + Z, a \rangle - c_t(a) \right] \geq u_t^0$$

type-by-type. Instead, we work with a *reduced-form* participation regularity condition and a conservative baseline set. One convenient sufficient condition for analysis is:

$$\Pr(p_t = 1 \mid b_t) \;\geq\; \pi_{\min} > 0 \quad \text{for all } t, \tag{26}$$

which says that the induced participation rate is uniformly bounded away from zero along the policy path. Operationally, (26) can be supported by constructing $\mathcal{B}_{\text{safe}}$ as a neighborhood of historically "high-entry" baselines, by applying guardrails (e.g., coordinatewise lower bounds on b_t), or by using a fallback mixture

$$b_t = (1 - \alpha_t)\, b^{\text{cons}} + \alpha_t\, \Pi_{\mathcal{B}}\big(\phi(\hat{\theta}_t)\big), \qquad \alpha_t \uparrow 1,$$

where b^{cons} is a conservative baseline known to yield acceptable participation. Importantly, the post-participation perturbation Z_t does not itself alter entry in our timing, so "safe exploration" is feasible: we can randomize to identify $\theta^\star$ without directly endangering participation.

Estimation error under selection: self-normalized martingale control. Define the selected-sample moment noise $\Gamma_t = y_t - \langle \theta^\star, x_t \rangle$ on participating rounds. Consider the matrix and vector processes

$$S_t \;\equiv\; \sum_{s \in \mathcal{T}_{1,t-1}} Z_s x_s^\top, \qquad g_t \;\equiv\; \sum_{s \in \mathcal{T}_{1,t-1}} Z_s \Gamma_s,$$

so that $\hat{\theta}_t - \theta^\star = (S_t + \lambda I)^{-1} g_t$ by construction. The key probabilistic fact is that, conditional on the filtration and on $p_s = 1$, the sequence $(Z_s \Gamma_s)_{s \in \mathcal{T}_{1,t-1}}$ is a martingale difference array: Z_s is mean-zero and independent of Γ_s even in the selected sample. This is precisely where post-participation randomization pays off; it restores orthogonality without requiring any assumptions about the correlation between ω_t and entry.

Under standard boundedness/subgaussian conditions on Z_t and Γ_t (conditional on $\mathcal{F}_t$ and $p_t = 1$), self-normalized inequalities imply a bound of the schematic form

$$\|\hat{\theta}_t - \theta^\star\|_2 \;\leq\; \tilde{O}\left(\frac{\sqrt{d \log(1/\delta)}}{\sigma_{\min}(S_t + \lambda I)} \right) \quad \text{uniformly over } t \leq T, \tag{27}$$

with probability at least $1 - \delta$. Selection enters (27) only through the growth and conditioning of S_t, which in turn are driven by (i) how many rounds participate (via $|\mathcal{T}_{1,t-1}|$) and (ii) how strongly actions load on the randomized directions (via the cross-moments between Z_t and x_t in the participating sample). In particular, if (26) holds and Σ_Z is well-conditioned, then $\sigma_{\min}(S_t)$ typically grows on the order of $|\mathcal{T}_{1,t-1}|$, yielding the familiar $1/\sqrt{|\mathcal{T}_{1,t-1}|}$ shrinkage in estimation error.

From estimation error to regret: stability of the baseline map. To translate (27) into regret guarantees, we impose a local regularity condition on the baseline choice problem. Let $\bar{\mu}(b;\theta)$ denote a smooth surrogate for the platform's expected payoff when the outcome model parameter is θ (for instance, the platform's expected value of induced effort net of payments, under the model-implied best response). We assume:

1. (*Local strong concavity*) For θ in a neighborhood of $\theta^\star$, the function $b \mapsto \bar{\mu}(b;\theta)$ is α-strongly concave on $\mathcal{B}_{\text{safe}}$.

2. (*Lipschitz dependence*) The gradient $\nabla_b \bar{\mu}(b;\theta)$ is L-Lipschitz in θ, uniformly over $b \in \mathcal{B}_{\text{safe}}$.

3. (*Plug-in optimality*) The update map ϕ in (25) approximates an optimizer of $b \mapsto \bar{\mu}(b;\hat{\theta}_t)$ up to a controlled error, or is itself the exact optimizer when tractable.

These conditions are standard in online estimation–optimization couplings: they state that the baseline problem is well-behaved and does not amplify small parameter errors into large contract mistakes. Under them, one obtains a stability inequality of the form

$$\|b_t - b^\star\|_2 \;\le\; C_b \, \|\hat{\theta}_t - \theta^\star\|_2, \tag{28}$$

for a constant C_b depending on (α, L) and on the Lipschitz properties of ϕ and projection. Combining (28) with a second-order expansion of $\mu_t(b)$ around $b^\star$ (using concavity) yields one-step regret bounded by a quadratic in $\|b_t - b^\star\|_2$, hence ultimately controlled by $\|\hat{\theta}_t - \theta^\star\|_2^2$.

A representative regret bound and its interpretation. Putting the pieces together, we obtain a regret guarantee of the familiar "parametric" flavor, but with two participation-sensitive modifiers. Under (26), the moment validity conditions for Z_t, and the stability assumptions above, a typical bound is

$$\text{Reg}(T) \;\le\; \tilde{O}\!\left(\frac{d}{\pi_{\min}}\sqrt{T}\right), \tag{29}$$

where logarithmic factors depend on confidence and tail parameters, and the constant depends on curvature and the conditioning of Σ_Z as mediated by behavior. The key economic content of (29) is not the specific $\sqrt{T}$ rate—which mirrors standard stochastic online learning—but rather the channels through which selection enters:

- *Data scarcity:* a smaller participation rate effectively reduces the sample size of valid moments, slowing the decay of $\|\hat{\theta}_t - \theta^\star\|$ and inflating regret.

- *Behavioral relevance:* even if $\Sigma_Z \succ 0$ by design, weak responsiveness of effort in certain dimensions can make the effective first stage ill-conditioned within $\mathcal{T}_1$, degrading both estimation and regret.

- *Policy-induced composition:* the baseline affects who shows up, which changes the distribution of costs and thus the mapping from incentives to actions. Our approach does not require this composition to be stable; it requires only that the post-entry randomization remains orthogonal to the residual, and that participation does not collapse.

Design implications and limitations. The preceding analysis highlights a practical design principle: we can decouple *validity* from *relevance*. Validity is guaranteed by timing—randomize after entry—whereas relevance must be engineered by choosing the covariance (and scale) of Z_t so that, among participants, Z_t generates sufficiently rich variation in x_t. This is precisely why instrument-strength diagnostics computed on the selected sample are not merely inferential conveniences but genuine control variables in the learning loop.

At the same time, we emphasize what this framework does *not* solve. First, participation feasibility is modeled through reduced-form stability conditions (such as (26)) or conservative safe sets; a fully structural treatment would require learning about the joint distribution of (c_t, u_t^0) and solving a dynamic mechanism design problem. Second, our regret benchmark is against a fixed baseline; when the environment is nonstationary, one may prefer adaptive benchmarks, at the cost of additional assumptions. Third, risk aversion or downside constraints would make the distributional shape of Z_t (not only its variance) central for feasibility; heavy tails can generate rare but severe realizations that reduce entry in practice even though entry is ex ante. These considerations motivate the extensions we discuss next.

5.6 Extensions: covariates, fairness, competition, delays, and nonstationarity

The analysis above isolates a simple but powerful idea: by randomizing *after* participation, we obtain moment conditions that remain valid in the selected sample. In practice, however, platforms rarely face a homogeneous stream of agents, immediate outcomes, or a monopolistic environment. We therefore briefly sketch several extensions that preserve the same logical separation between (i) economic forces that govern entry and effort and (ii) statistical validity of the instrument, while clarifying where new modeling work is genuinely required.

Observed covariates and contextual contracts. Suppose that, before posting the baseline, the platform observes a vector of context variables

$w_t \in \mathbb{R}^m$ (e.g., market conditions, seller experience, product category, time of day). A natural generalization is to allow both the contract and the outcome model to depend on w_t. One convenient formulation is a linear-in-features outcome model

$$y_t \;=\; \langle \theta^\star, a_t \rangle + \langle \psi^\star, w_t \rangle + \omega_t + \eta_t,$$

or, more flexibly, $\theta^\star$ itself may vary with context via a known feature map $\varphi(w_t)$, e.g.,

$$y_t \;=\; \big\langle \Theta^\star \varphi(w_t),\, a_t \big\rangle + \omega_t + \eta_t,$$

where $\Theta^\star$ is an unknown matrix. On the contract side, we can allow $b_t = b(w_t)$ for some policy class, with the perturbation still realized post-entry: $\beta_t = b(w_t) + Z_t$.

The key observation is that the selection-robust moment condition continues to hold *conditionally on covariates*:

$$\mathbb{E}\big[Z_t \big(y_t - \langle \theta^\star, x_t \rangle - \langle \psi^\star, w_t \rangle \big) \,\big|\, p_t = 1,\, w_t \big] = 0,$$

under the same timing and independence assumptions on Z_t. This immediately suggests a standard strategy: instrument not only with Z_t, but with interactions $Z_t \otimes \varphi(w_t)$ to identify heterogeneous task values. Estimation becomes a selected-sample analog of contextual IV, with the usual rank condition replaced by nonsingularity of the conditional (or feature-augmented) cross-moment matrix $\mathbb{E}[(Z_t \otimes \varphi(w_t))\, x_t^\top \mid p_t = 1]$. Economically, contextual baselines are attractive precisely because they can keep participation feasible by tailoring incentives to the observable environment; statistically, they can also strengthen the first stage by targeting contexts where effort responds more elastically.

A limitation is that if w_t affects not only the level of y_t but also the distribution of unobserved ω_t in ways correlated with the policy, then the platform may wish to explicitly condition its safe-set construction on w_t (e.g., ensuring $\Pr(p_t = 1 \mid w_t, b_t) \geq \pi_{\min}(w_t)$). This is not a failure of validity—the instrument remains orthogonal—but rather a practical requirement to avoid "contextual data deserts" in which certain covariate regions never generate participants.

Group fairness and constrained contract updates. Platforms may face normative or regulatory constraints that require contracts to satisfy fairness criteria across protected groups. Let $g_t \in \{1, \ldots, G\}$ denote an observed group label (or a coarse proxy), and consider fairness constraints that operate either on *participation* (e.g., demographic parity of entry) or on *treatment intensity* (e.g., bounds on differences in expected payments).

Our framework accommodates such constraints most naturally at the baseline-update stage. For example, if the policy class is group-conditional $b_t = b(g_t)$, one can impose constraints of the form

$$\big|\, \Pr(p_t = 1 \mid g_t = g, b(g)) - \Pr(p_t = 1 \mid g_t = g', b(g')) \big| \leq \Delta,$$

or, more conservatively, group-wise feasibility bounds $\Pr(p_t = 1 \mid g_t = g, b(g)) \geq \pi_{\min,g}$. Alternatively, one may require that the baseline itself satisfy Lipschitz or bounded-disparity constraints $\|b(g) - b(g')\|_2 \leq \delta$.

Crucially, the post-participation instrument Z_t remains valid *within each group*:

$$\mathbb{E}\big[Z_t\big(y_t - \langle \theta^\star, x_t \rangle\big) \,\big|\, p_t = 1,\, g_t = g\big] = 0,$$

so one can estimate either a common $\theta^\star$ pooling all groups, or group-specific parameters $\theta_g^\star$ when heterogeneity is substantively important. The policy implication is that fairness constraints need not force the platform to abandon identification; instead, they change the feasible set over which the plug-in optimizer operates and can reduce effective sample size for some groups. In turn, regret guarantees become group-sensitive, scaling with the smallest group participation rate and with the weakest within-group first stage. This highlights a tension that is easy to miss in purely static analyses: fairness constraints can be binding precisely in the regions where learning is hardest, so auditing and monitoring should include instrument-strength diagnostics at the group level (e.g., $\sigma_{\min}(\mathbb{E}[Zx^\top \mid p = 1, g])$) rather than only aggregate metrics.

Multi-homing and platform competition (reduced form). Many marketplaces face multi-homing: agents can participate on multiple platforms, or choose among competing contracts posted elsewhere. A reduced-form way to incorporate this is to reinterpret the outside option u_t^0 as an equilibrium value that depends on competitors' terms, macro conditions, and agent-specific switching costs. Then $p_t = 1$ becomes a market-share event rather than a pure participation decision.

Two issues arise. First, the baseline b_t may now affect not only selection on unobservables but also the competitive equilibrium, potentially changing the distribution of arriving types (e.g., high-quality agents sort to the platform offering better incentives). Second, competitors may respond strategically over time, inducing correlation between the platform's policy and the environment.

For identification of $\theta^\star$ *from within-platform data*, the timing logic still helps: conditional on the event that an agent chose this platform (i.e., conditional on $p_t = 1$ as defined by "arrived and accepted here"), a post-entry perturbation Z_t that is independent of the agent and contemporaneous shocks continues to satisfy

$$\mathbb{E}[Z_t(y_t - \langle \theta^\star, x_t \rangle) \mid p_t = 1] = 0.$$

Thus, even if competition makes selection more severe, it does not by itself invalidate the instrument. What it does change is the feasibility problem: the safe set $\mathcal{B}_{\text{safe}}$ must now ensure a lower bound on equilibrium participation/share, and this may require explicit guardrails tied to observable

competitor signals (when available). Moreover, the platform's objective may become explicitly game-theoretic, so regret relative to a fixed baseline in hindsight may be less meaningful than regret relative to an equilibrium benchmark or to a constrained best-response set. We view this as an important direction for future work: extending "safe randomization" to strategic settings where the environment endogenously reacts to the baseline, while preserving post-entry orthogonality.

Delayed outcomes and asynchronous updating. In many applications, x_t is observed immediately (telemetry, intermediate outputs) but y_t arrives with delay (chargebacks, retention, long-run quality). Let y_t be observed only at time $t+\ell_t$, with possibly random lag ℓ_t. The estimator (24) can be adapted by updating only when the corresponding y_s becomes available:

$$\hat{\theta}_{t+1} = \left(\sum_{s \in \mathcal{O}_t} Z_s x_s^\top + \lambda I \right)^{-1} \left(\sum_{s \in \mathcal{O}_t} Z_s y_s \right),$$

where $\mathcal{O}_t$ indexes participating rounds whose outcomes have arrived by time t. Validity is unchanged because it is a *within-round* statement: Z_s remains independent of the residual in round s, regardless of when that residual is observed.

What does change is the online control problem: the baseline updates become "stale" because they are driven by delayed information. Under bounded delays, standard arguments for online learning with delayed feedback suggest regret inflation that depends on the maximum delay, reflecting the fact that the policy makes more decisions before incorporating new data. Substantively, delayed outcomes strengthen the case for maintaining persistent (but safe) randomization: without it, long feedback loops can easily lead to premature lock-in to poorly identified baselines. In settings with very long delays, it may be valuable to use x_t-only diagnostics (e.g., first-stage strength, compliance) to adapt the scale σ_t in real time, even before y_t arrives.

Nonstationarity: drifting values and evolving populations. Finally, both $\theta^\star$ and the type distribution may evolve over time. Some changes are predictable (seasonality) and can be absorbed into covariates; others reflect genuine drift (changing user tastes, policy shocks, product-market fit). A pragmatic extension is to replace the full-sample estimator with a discounted or sliding-window version, for example,

$$\hat{\theta}_t = \left(\sum_{s \in \mathcal{T}_{1,t-1}} \rho^{t-1-s} Z_s x_s^\top + \lambda I \right)^{-1} \left(\sum_{s \in \mathcal{T}_{1,t-1}} \rho^{t-1-s} Z_s y_s \right), \qquad \rho \in (0,1),$$

or with a window of the last W participating observations. Here, post-participation randomization again ensures that the *moment condition is correct at each date* (relative to the contemporaneous $\theta_t^\star$), while the discounting controls the bias–variance tradeoff induced by drift.

The economic message is that nonstationarity primarily re-enters through relevance and feasibility. Drift can move the system into regions where participation is fragile or where responsiveness to incentives weakens, thereby shrinking the effective information rate even if the formal IV moment remains unbiased. This suggests that safe-set design and instrument scaling should be treated as adaptive control knobs, not static assumptions. It also suggests more demanding benchmarks: rather than competing with the best fixed baseline, one may compare to a slowly varying sequence of baselines, paying a variation budget. Establishing sharp regret bounds in such environments is feasible but requires explicit drift controls and is beyond our scope here.

Summary. Across these extensions, the common theme is modularity. Observed covariates and fairness constraints primarily alter how we map estimates into baselines and how we define feasibility; competition alters what participation means and may change the appropriate benchmark; delays alter the information pattern but not instrument validity; and nonstationarity alters the target and thus the estimator design. In all cases, the central "safe randomization" insight remains: when perturbations are realized after entry and are exogenous, selection can distort who we observe without corrupting the orthogonality of the instrument in the observed sample. This prepares the ground for our discussion of what the approach implies for platform implementation, auditing, and policy.

5.7 Discussion: platform implementation, audit design, policy implications, and open problems

The preceding sections emphasize a conceptual separation: participation and effort are governed by economic incentives and selection, while identification can be recovered by a carefully timed source of exogenous variation. This separation is attractive in platforms precisely because it maps onto how systems are built. Participation decisions are typically made at an "offer layer" (what terms are posted, what sellers see, whether a worker accepts a job), whereas effort and performance are realized only after a match occurs. Our view is that post-participation randomization is best interpreted as an engineering principle for experimentation under endogenous observation: we should randomize *only* along dimensions that are realized after the platform has committed to observing the relevant data.

How a platform would implement post-participation randomization. Operationally, the platform needs to (i) publish a baseline slope vector b_t and a distribution for Z_t, (ii) draw Z_t only after an agent has accepted/entered, and (iii) record the realized $\beta_t = b_t + Z_t$ alongside x_t and y_t for estimation. A simple design is to take

$$Z_t = \sigma_t \, \xi_t, \qquad \mathbb{E}[\xi_t] = 0, \qquad \mathbb{E}[\xi_t \xi_t^\top] = I,$$

with $\sigma_t \geq 0$ chosen by the platform (possibly time-varying) and ξ_t generated by a centralized randomness service. The requirement that Z_t be realized *after* participation is not merely conceptual: it is a product requirement. In particular, the user interface, API responses, and any pre-acceptance previews must depend only on b_t and on the *distributional description* of Z_t (e.g., "your per-unit bonuses may vary slightly around the posted rates") rather than on the realized draw. When this timing is respected, the event $p_t = 1$ cannot mechanically encode information about Z_t, which is the core reason the selected-sample moment remains valid.

Two practical constraints often matter. First, platforms typically require nonnegativity or other monotonicity in incentives (e.g., $b_t \in [0,1]^d$). This can be handled by choosing a perturbation distribution supported on a set that preserves feasibility, such as a truncated Gaussian, a Rademacher design with small amplitude, or a "reflecting" scheme that redraws until $\beta_t \in \mathcal{B}$. Second, payments implied by $\langle \beta_t, x_t \rangle$ may be subject to budgets, caps, or risk controls. These constraints naturally enter through $\mathcal{B}$ and through an explicit experimentation cost term (as in $-\kappa \|Z_t\|_2^2$), which provides a direct knob to trade off short-run payment volatility against long-run learning.

Choosing the scale and shape of randomization. From an implementation standpoint, the most important design choice is not whether to randomize but *how much* and in *which directions*. The variance Σ_Z governs instrument strength through the cross-moment $\mathbb{E}[Z_t x_t^\top \mid p_t = 1]$ and therefore directly affects statistical power. At the same time, large perturbations can create undesirable variability in realized incentives, which may harm trust, increase churn, or distort effort away from what the baseline would have elicited.

A useful way to think about tuning is to treat Σ_Z as an exploration budget allocated across tasks. If some coordinates of effort are already strongly responsive to incentives (high "compliance"), then randomization in those directions yields high first-stage strength at low variance; conversely, if certain tasks are inelastic, randomization may need to be larger to identify their value, or the platform may decide that those dimensions are effectively unlearnable under acceptable perturbations. In practice, we recommend that platforms monitor a rolling estimate of the smallest singular value of the empirical first-stage matrix $Z_{\mathcal{T}_1}^\top X_{\mathcal{T}_1}$ (or its regularized analog) and adapt σ_t

to maintain it above a minimum threshold. This reframes weak-instrument concerns as a control problem: we are not passively accepting weak relevance; we are actively maintaining it subject to safety and user-experience constraints.

Data logging and "separation of duties" in the system. Because the identifying variation comes from a timing claim, the most fragile failure modes are engineering failures. A robust implementation therefore benefits from separation of duties: a service that decides eligibility/participation should not have access to the realized Z_t, and a service that draws Z_t should be triggered only after participation is irrevocably recorded. Moreover, the platform should log (at minimum) $(t, b_t, Z_t, \beta_t, p_t, x_t, y_t)$ with immutable timestamps and stable identifiers, so that later audits can verify that the realized perturbations were not leaked or retroactively modified.

We also emphasize that, in many products, the "signal" x_t is itself a derived telemetry object rather than raw behavior. When x_t is constructed by downstream pipelines, it is important that its construction be invariant to Z_t except through the agent's behavior. Any direct dependence of measurement on the instrument (e.g., changing logging intensity when incentives are high) can reintroduce endogeneity in a way that is subtle but empirically consequential. Treating the instrument as a protected variable in the data schema—available for estimation but not for measurement logic—is therefore a practical safeguard.

Estimation and decision pipelines in a selected sample. A virtue of the selected-sample IV moment is that it matches how platform data are naturally generated: x_t and y_t exist only when $p_t = 1$. Estimation can thus be implemented as a streaming two-stage least squares or GMM routine that updates only on participating rounds, with regularization to handle transitory weak relevance:

$$\hat{\theta}_t = \left(\sum_{s \le t-1 : p_s = 1} Z_s x_s^\top + \lambda I \right)^{-1} \left(\sum_{s \le t-1 : p_s = 1} Z_s y_s \right).$$

The baseline-update policy then maps $\hat{\theta}_t$ into $b_t \in \mathcal{B}$, potentially via a smooth optimizer or via a conservative projection rule. We find it helpful to make this mapping explicit and auditable (e.g., "baseline is the solution to a constrained surrogate problem with parameter $\hat{\theta}_t$"), because it clarifies how incentives will change as the estimate updates and allows internal stakeholders to reason about stability.

Audit design: validating the timing assumptions and diagnosing strength. In applied settings, the right question is not whether an as-

sumption is philosophically plausible, but whether it can be *audited*. The core assumptions behind instrument validity are (a) post-participation realization and (b) statistical independence of Z_t from unobservables and noises. Both admit concrete tests and monitoring.

First, one can run "balance" checks that should hold mechanically if timing is correct: since Z_t is drawn after p_t is realized, we should have $\Pr(p_t = 1 \mid Z_t) = \Pr(p_t = 1)$ up to sampling error, and likewise $\mathbb{E}[Z_t \mid p_t = 1] = 0$. While these are not sufficient to guarantee full independence, they are sensitive to common implementation bugs (e.g., accidentally drawing Z_t before acceptance and using it in ranking/eligibility). Second, instrument strength diagnostics should be mandatory: the platform should track the empirical covariance between Z_t and x_t among participants, report weak-instrument flags, and condition confidence intervals on robust (heteroskedasticity- and autocorrelation-robust) estimators when appropriate. Third, when fairness or group constraints are present, these diagnostics should be computed within groups, because weak identification can be concentrated precisely where participation is low.

Beyond these mechanics, we advocate for "placebo" outcome checks as an ongoing audit tool. If there are outcomes that, by design, cannot respond to contemporaneous effort (or cannot respond within the delay window), then they should have zero reduced-form relationship with Z_t. A statistically significant relationship is then an actionable signal of leakage, measurement dependence, or correlated shocks, even if the main outcome regression appears well-behaved.

Implications for policy and platform governance. Post-participation randomization changes how we should think about experimentation in markets with selection. Standard A/B testing logic implicitly assumes that assignment is orthogonal to what is observed; here, what is observed depends on participation, so naive experiments can generate misleading conclusions even when assignment is randomized at the offer stage. Our framework suggests a governance principle: to credibly learn about marginal returns to effort, platforms should randomize *incentives conditional on entry*, not just offers that affect entry.

For regulators and auditors, this yields a practical standard for acceptable experimentation. The platform can commit to (i) a publicly documented perturbation distribution, (ii) a maximum variance (protecting participants from excessive volatility), and (iii) monitoring of participation and payment impacts. This is analogous to how clinical trials codify dose randomization within a safe range. Importantly, the approach also clarifies what cannot be inferred from platform data alone: if certain populations never participate under any feasible baseline, then no amount of post-entry randomization can identify their counterfactual outcomes. In that sense, the method is not a

substitute for access policies; it is a complement that makes inference within the observed market more credible.

Limitations and open problems. Several limitations deserve emphasis. First, we have treated each round as a one-shot interaction, but many platforms face dynamic incentives and repeated participation by the same agents. If agents learn over time about the distribution of Z_t and update beliefs about future baselines, participation and effort may become forward-looking, and the effective timing assumptions can blur (e.g., an agent may anticipate that accepting today affects future treatment). Extending selection-robust IV to dynamic contracts is feasible but requires explicit modeling of state and beliefs, and may call for randomized policies that are conditionally independent given states.

Second, we have assumed that Z_t is independent of the agent type and contemporaneous shocks. In practice, correlated shocks can arise from shared infrastructure (e.g., outages) or from targeting logic that inadvertently correlates Z_t with contexts that also affect outcomes. This points to a design desideratum: generate Z_t from a centralized, context-agnostic random seed, and treat any context-dependent scaling σ_t as part of the baseline decision b_t that is itself predictable from $\mathcal{F}_t$. When such predictability fails, one may need to condition on richer information sets or adopt randomization at higher granularity.

Third, the outcome model itself may be misspecified. If y_t depends nonlinearly on a_t, or if there are complementarities across tasks, the linear IV estimator targets a local linear approximation rather than a structural primitive. This is not necessarily a defect—platform decisions often only require marginal values—but it does affect interpretation and optimal contract design. A promising direction is to combine post-entry randomization with flexible outcome models (e.g., series or machine learning) while retaining orthogonality through moment restrictions, essentially moving from linear IV to orthogonal score estimation in a selected sample.

Finally, there is an unresolved design question at the heart of implementation: how should a platform optimally choose Σ_Z jointly with the baseline policy to maximize long-run welfare subject to volatility, fairness, and participation constraints? Our regret discussion provides one tractable benchmark, but real platforms optimize multi-objective criteria and face organizational constraints (communication, user trust, regulatory scrutiny) that are not naturally captured by a single payoff function. Developing principled "experimentation budgets" that translate these constraints into transparent bounds on perturbations, while preserving identification, is an important open problem for both economics and platform science.

Taken together, these considerations reinforce the main message. Safe randomization is not merely a statistical trick; it is a disciplined way to

align mechanism design with how data are generated under selection. When implemented with explicit timing guarantees and audited for strength and leakage, it offers a credible path to learning causal task values in environments where naive instruments and naive experiments fail.