Auditing Algorithmic Pay: Testable Signatures of Free-Fall Dynamic Contracting against Learning Agents

Liz Lemma   Future Detective

January 16, 2026

Abstract

Algorithmic management systems increasingly adjust compensation rules over time while workers (or AI assistants acting for workers) learn how to respond. Recent theory shows that against mean-based no-regret learners, optimal dynamic contracting can take a simple "free-fall" form: offer a high incentive briefly, then drop incentives to zero, inducing the learner to cascade through actions while the principal continues earning reward at low cost. This paper turns that mechanism into an audit problem. We propose a clean binary-outcome model with linear contracts and latent actions, where an external auditor observes posted contract rates and realized outcomes across many agents. Using the continuous-time characterization of mean-based learning (best response to the historical average contract), we derive moment inequalities and shape restrictions that must hold if observed data are generated by free-fall-like manipulation. We provide a consistent specification test and a structural estimator that reconstructs implied breakpoints (indifference points) and quantifies a lower bound on surplus transfer relative to a counterfactual static benchmark. The contribution is methodological: converting dynamic-contracting-against-learning theory into falsifiable empirical restrictions and practical compliance metrics for 2026-era platforms.

Table of Contents

3. 3. Model: binary outcomes, linear contracts, latent actions, mean-based learning approximation, multi-agent panel data, churn extension (optional).

4. 4. Continuous-time restrictions as testable implications: best-response-to-historical-average constraints; characterization of free-fall signature (hyperbolic average contract, breakpoint hitting times).

5. 5. Identification strategy: from outcomes to latent action segments (change-point + separation), from contracts to historical averages, mapping to implied breakpoints; what is point-identified vs set-identified.

6. 6. Specification tests: moment-inequality test for mean-based compatibility; sharper test for free-fall (one-switch) structure; size/power discussion and finite-sample calibration.

7. 7. Estimation of manipulation magnitude: estimate implied breakpoint sequence and bound principal surplus transfer vs best static benchmark; uncertainty quantification.

8. 8. Simulation study: synthetic data generated from the source model (mean-based learners + free-fall principal), robustness to exploration, noise, heterogeneity, and partial observability.

9. 9. Empirical application plan (data permitting): what data are needed, plausible proxies, and how to implement the audit; alternatively, a detailed 'audit protocol' if proprietary data are unavailable.

10. 10. Discussion and extensions: multi-KPI contracts, stochastic stopping/churn, multi-agent interactions, and regulatory implications.

1 1. Introduction: algorithmic pay, learning agents, and why 'dynamic manipulation' should be auditable; summary of contributions and the free-fall mechanism.

Algorithmic compensation systems are increasingly deployed in environments where the platform observes rich performance metrics but neither the regulator nor the worker fully understands how incentives evolve over time. In such settings, a linear pay-for-success rule can be posted and updated automatically, and workers respond not by solving a static optimization problem once and for all, but by learning which actions "pay off" through experience. This combination of automated contracting and adaptive behavior creates a natural concern for auditors: a platform may be able to extract surplus through dynamic paths of incentives that look innocuous when viewed period by period, yet systematically shape behavior over the longer run. Our objective is to make this type of "dynamic manipulation" empirically auditable using only the information that is typically available in practice—the posted contract sequence and realized outcomes.

The central friction is informational. The auditor does not observe effort or the worker's action, and usually cannot credibly measure the underlying cost of effort. At the same time, the platform can commit to (or at least implement) a deterministic sequence of contract terms $\{\alpha_t\}_{t=1}^{T}$, where the worker is paid $\alpha_t \in [0, 1]$ upon success and 0 upon failure. In many applications, these contract terms are themselves a function of an internal algorithm, potentially responsive to earlier outcomes, retention, or growth targets. A naive audit that treats each round as independent will miss the key channel: when workers use a mean-based learning algorithm, their behavior can be well-approximated as a best response not to the current contract α_t, but to a smoothed statistic such as the historical average $\bar{\alpha}_t = \frac{1}{t}\sum_{s=1}^{t} \alpha_s$. Hence a platform can influence current effort by shaping the entire history of incentives.

We study a parsimonious principal–agent environment designed to isolate this channel. Agents repeatedly choose among finitely many ordered actions $a \in \{1, \ldots, n\}$, where higher actions have higher cost c_a and strictly higher success probability q_a. Outcomes $y_{i,t} \in \{0, 1\}$ are observed, and the platform's per-success revenue is normalized to one. Under a linear contract α_t, the agent's expected utility from action a is $u^A(\alpha_t, a) = \alpha_t q_a - c_a$, while the principal's expected payoff is $u^P(\alpha_t, a) = (1 - \alpha_t)q_a$. The key behavioral assumption is not full rationality or perfect foresight, but mean-based no-regret learning: over time, each agent shifts probability mass toward actions that have performed well in terms of realized average utility. A convenient continuous-time proxy of this learning dynamic implies that play concentrates on a best response to $\bar{\alpha}_t$ for most t, up to an algorithmic slack

$\gamma(T) = o(1)$.

This learning structure makes a particular form of dynamic manipulation both feasible and, crucially, testable. We focus on a "free-fall" contract path: the platform offers a constant incentive α_0 for an initial phase $t \leq t_0$, then drops the incentive to zero thereafter, i.e.,

$$\alpha_t = \alpha_0 \mathbf{1}\{t \leq t_0\}.$$

Even though $\alpha_t = 0$ for $t > t_0$, the historical average declines only gradually,

$$\bar{\alpha}_t = \begin{cases} \alpha_0, & t \leq t_0, \\ \alpha_0 t_0/t, & t > t_0, \end{cases}$$

so a mean-based learner continues to behave as if incentives remain positive for some time. Intuitively, the platform "front-loads" incentives to move the worker into a high-success action and then "coasts" on the worker's inertia as the average decays. Eventually, as $\bar{\alpha}_t$ falls, the worker becomes indifferent between adjacent actions. Denoting the indifference breakpoint between actions $a-1$ and a by

$$\alpha_{a-1,a} = \frac{c_a - c_{a-1}}{q_a - q_{a-1}},$$

the learning dynamics imply that action downgrades occur when $\bar{\alpha}_t$ passes these thresholds. Under free-fall, this generates a rigid, approximately hyperbolic timing pattern: the time at which the population switches from a to $a-1$ must satisfy $\bar{\alpha}_{t_a} \approx \alpha_{a-1,a}$, and hence $t_a \approx (\alpha_0 t_0)/\alpha_{a-1,a}$. This inverse proportionality is the mechanism's statistical fingerprint.

Our first contribution is to translate this fingerprint into restrictions that an auditor can test using only contracts and outcomes. Because actions are latent, we use the cross-sectional success rate $\hat{s}_t = \frac{1}{N}\sum_{i=1}^{N} y_{i,t}$ as a sufficient statistic for effort in large populations. The separation condition $\Delta_q = \min_{a\geq 2}(q_a - q_{a-1}) > 0$ implies that distinct actions induce distinct success rates, so abrupt changes in $\hat{s}_t$ can be interpreted as transitions in the dominant action. Combining this outcome segmentation with the mean-based best-response characterization yields a finite set of moment inequalities: within any time segment where outcomes are consistent with action a, the historical average $\bar{\alpha}_t$ must lie in the interval of contracts that make a optimal, up to slack of order $O(\gamma(T))$ and sampling error $O_P(1/\sqrt{N})$. These inequalities provide necessary conditions for compatibility with free-fall manipulation without requiring the auditor to observe costs.

Our second contribution is a specification test that distinguishes free-fall-compatible manipulation from benign contracting paths that do not generate the hyperbolic "boundary-hitting" sequence. The logic is simple: under free-fall, multiple action drops must align with a single pair of parameters (α_0, t_0) through the mapping $\bar{\alpha}_t = \alpha_0 t_0/t$ after t_0. A static contract, or a piecewise-constant rule without a free-fall phase, can produce changes in outcomes, but generically will not reproduce the implied relationship between change-point times and breakpoints across several transitions. We formalize this contrast using moment-inequality inference with nuisance parameters $\{q_a\}$ that are only set-identified, thereby controlling size under the null while retaining power against structured dynamic manipulation.

Our third contribution is constructive: when the data are compatible with free-fall, we show how an auditor can recover the implied breakpoint sequence and compute a conservative lower bound on the platform's incremental payoff relative to the best static linear contract. The economic content is that dynamic manipulation is not merely a qualitative concern; it can be quantified as surplus transfer. Using the observed α_t and $\hat{s}_t$, we can form an empirical estimate of the platform's realized payoff path and compare it to a static benchmark that optimizes $(1-\alpha)\hat{q}(\alpha)$ given the inferred mapping from incentives to success probabilities. Because costs are unobserved, we emphasize lower bounds that remain valid under set identification and learning slack.

Several limitations are worth stating at the outset. Our approach relies on the mean-based nature of learning and the existence of a stable ordering of success probabilities across actions; if outcomes are heavily nonstationary for reasons unrelated to effort, or if the population is highly heterogeneous in unmodeled ways, outcome segmentation becomes less informative. Likewise, a sufficiently sophisticated platform could randomize contracts or introduce additional state variables to obscure the historical-average channel. We view these caveats not as defects of the exercise, but as clarifying where auditing leverage comes from: the audit is powerful precisely when incentive paths are simple enough to be implementable at scale, and worker learning is structured enough to create predictable behavioral inertia.

2 Related Literature

Our analysis builds on several strands of work that study how incentives interact with boundedly rational or adaptive agents, and how such environments can be disciplined empirically when key primitives are unobserved. The common theme is that, once behavior is shaped by learning dynamics rather than static optimization, the relevant object for design and inference is not a one-shot best response but a path-dependent mapping from past incentives to current actions. This perspective is increasingly central in digital labor markets and platform settings, where contract terms are updated by algorithms and workers experiment, imitate, and gradually adjust effort in response to experienced payoffs.

Contracting with learning agents. A classic benchmark in contracting is the repeated moral hazard model with linear incentives and risk-neutral parties (?), where effort responds contemporaneously to current pay parameters. Our setting departs from this benchmark not by altering information or risk-sharing, but by relaxing the behavioral assumption that agents instantaneously play a best response to the current contract. Recent work in dynamic mechanism design and principal–agent problems has begun to incorporate learning or limited attention, emphasizing that principals may exploit inertia, habituation, or reference dependence to reduce payments while preserving output (??). A related literature studies agents who learn about the environment (or about their own productivity) over time, so that contracts influence both current actions and beliefs (??). We view mean-based no-regret learning as a complementary behavioral primitive: rather than modeling a particular bias, we impose a robust adaptive property (low regret) and exploit its implication that long-run play tracks a best response to a smoothed payoff signal.

Within the broader learning-in-games literature, no-regret dynamics are known to yield sharp predictions about which actions can persist (those near the top of the average-payoff distribution) and how quickly dominated actions are discarded (??). The "mean-based" refinement we rely on, which bounds play on actions that are far behind in cumulative utility, is particularly useful for identification because it converts a potentially complicated stochastic adjustment process into tractable inequality restrictions with explicit slack (?). Relative to models of fully strategic forward-looking agents, this approach trades off richness of equilibrium path predictions for robustness: many learning algorithms satisfy the same mean-based property, so the audit implications do not hinge on a specific parametric updating rule.

Econometrics for learning and adaptive behavior. Empirical work on learning typically confronts an identification challenge: the researcher observes actions and outcomes but not the counterfactual payoffs that drive updates. Structural approaches specify a parametric learning rule (e.g., reinforcement learning, belief learning) and estimate it by simulated method of moments or maximum likelihood (?). A separate line develops revealed-preference or regret-based methods that can test rationality or bound preferences using only observed play (??). Our approach is closer in spirit to the latter: we do not attempt to recover each agent's learning parameters, but instead exploit an implication of low regret—approximate optimality with respect to the historical payoff environment—to generate moment inequalities that are directly testable in panel outcome data.

Methodologically, our use of change-point structure connects to econometric work on structural breaks and segmentation in time series and panel data (?). The novelty is that the breakpoints we target are not exogenous regime shifts, but endogenous behavioral transitions induced by a deterministic incentive path filtered through learning. In large populations, cross-sectional averaging reduces idiosyncratic noise, making the success-rate series a convenient sufficient statistic for effort shifts. This observation aligns with a growing literature on "macro" inference from many similar units (workers, users, sellers) interacting with a common platform policy, where the econometric object is an aggregate response function to a posted rule (**?**).

Empirical mechanism design and dynamic incentives. Our audit problem is also related to empirical mechanism design, which aims to infer or evaluate mechanism performance when agents respond strategically to design parameters (**??**). A key lesson from this literature is that, even when primitives such as valuations or costs are unobserved, mechanisms often impose testable restrictions (monotonicity, envelope conditions, incentive compatibility inequalities) that can be leveraged for estimation and counterfactuals. We adopt this logic but in a dynamic environment where incentive compatibility is behavioral (no-regret/mean-based) rather than equilibrium-theoretic. The resulting restrictions take the form of breakpoint and boundary-hitting inequalities: instead of identifying a full type distribution, we identify (or set-identify) the contract thresholds at which behavior shifts. This is analogous to identification via kink points in static screening models, but with the additional structure that the relevant state variable is a running average of past incentives.

There is also a natural connection to the literature on dynamic pricing and experimentation by platforms, including bandit-style policies and reinforcement-learning-based decision rules. Much of that work focuses on how a decision-maker learns demand or treatment effects over time (**?**). In contrast, we emphasize the reverse channel: the platform may already know how incentives map to outcomes, but can nonetheless use time variation in incentives to exploit worker learning and extract rents. From an auditing perspective, this distinction matters because it suggests that observed non-stationarity in incentives is not automatically evidence of experimentation or optimization under uncertainty; it may reflect deliberate intertemporal manipulation of a predictable behavioral dynamic.

Platform audits, accountability, and regulatory practice. Finally, our contribution speaks to an emerging policy and measurement agenda on algorithmic accountability in labor and marketplace platforms. Regulators and civil-society auditors often observe the platform's posted rules (payment formulas, bonus schemes, deactivation thresholds) and the realized outcomes (earnings, completion rates, quality metrics), but lack access to internal state variables, worker effort, or the platform's objective function. This partial observability has motivated a growing empirical literature on external auditing of algorithmic systems, including black-box tests, field audits, and statistical detection of rule changes (**??**). Our framework contributes an economic structure to this toolkit: it identifies a concrete signature of a particular dynamic incentive manipulation (a front-loaded contract followed by a collapse) that can be assessed using only standard administrative traces.

We also see our approach as complementary to transparency mandates that require disclosure of contract terms but not necessarily the full algorithm. Even when the posted sequence $\{\alpha_t\}$ is observable, its behavioral impact can be opaque if agents learn from long-run averages or other smoothed signals. By translating a learning-based behavioral model into moment inequalities that the observed data must satisfy, we provide a way to separate "innocent" time variation from time variation that is structurally consistent with surplus extraction via inertia. At the same time, we acknowledge a limitation that is shared by most audit methodologies: sufficiently sophisticated platforms can add randomness, personalization, or multidimensional scoring rules to blur the mapping from posted incentives to effort, weakening the power of any test based on aggregate outcomes. In this sense, our results clarify both the promise and the boundaries of empirical accountability: auditing leverage is greatest when incentive schemes are simple, scalable, and therefore constrained in the kinds of dynamic patterns they can generate.

3 Model

We study a repeated principal–agent environment in which the platform (the principal) chooses a sequence of linear incentive parameters, while a large population of agents adapt their actions over time using a mean-based no-regret learning rule. The key empirical feature is that agents' actions are unobserved, so an auditor must reason from posted contracts and realized outcomes alone. We keep the primitives deliberately sparse—binary outcomes, a finite ordered action set, and linear payments—both because these features match many platform pay schemes (bonuses for completion/acceptance/quality) and because they yield sharp restrictions once combined with learning.

Agents, actions, and outcomes. There are N ex ante identical agents indexed by $i \in \{1, \ldots, N\}$, and the interaction unfolds over rounds $t \in \{1, \ldots, T\}$. In each round, agent i privately chooses a latent action $a_{i,t} \in \{1, \ldots, n\}$. Action a should be interpreted as a coarse effort or compliance level: higher actions are more costly but make success more likely. Formally, each action a has a (time-invariant) cost c_a and success probability q_a, with strict ordering

$$0 = c_1 < c_2 < \cdots < c_n, \qquad 0 = q_1 < q_2 < \cdots < q_n.$$

Conditional on $a_{i,t} = a$, an observable binary outcome $y_{i,t} \in \{0, 1\}$ is realized according to

$$y_{i,t} \sim \text{Bernoulli}(q_a),$$

independently across agents and (conditional on actions) across time. The strict ordering of $\{q_a\}$ is a separation condition: it rules out observational equivalence across adjacent actions and will allow us to map persistent changes in observed success rates into changes in latent action mixtures in large populations.

Contracts and per-round payoffs. Before agents act in round t, the platform posts a linear contract parameter $\alpha_t \in [0, 1]$ that is common across agents. The realized wage is paid only upon success:

$$w_{i,t} = \alpha_t y_{i,t}.$$

We normalize the platform's gross benefit from success to $r = 1$ without loss of generality, so that α_t has the interpretation of "the share of output paid to the agent upon success." Given action $a_{i,t}$, the per-round monetary payoffs are

$$\pi^P_{i,t} = (1 - \alpha_t) y_{i,t}, \qquad \pi^A_{i,t} = \alpha_t y_{i,t} - c_{a_{i,t}}.$$

Taking expectations conditional on action a yields the expected utilities

$$u^P(\alpha_t, a) = (1 - \alpha_t) q_a, \qquad u^A(\alpha_t, a) = \alpha_t q_a - c_a.$$

Two implications are worth emphasizing. First, the agent's expected utility is affine in α_t, with slope q_a; this single-index structure is what makes historical averaging of incentives central under cumulative-payoff-based learning rules. Second, the platform's per-round payoff is increasing in success but decreasing in α_t, so the platform faces a standard tradeoff between inducing higher q_a and paying a larger success-contingent rent.

Timing and information. Each round t proceeds as follows: (i) the platform posts α_t; (ii) each active agent i selects $a_{i,t}$; (iii) outcomes $y_{i,t}$ are realized; (iv) wages $w_{i,t} = \alpha_t y_{i,t}$ are paid. We assume the auditor observes the full contract path $\{\alpha_t\}_{t \leq T}$ and the realized panel $\{(y_{i,t}, w_{i,t})\}$ for the agents who are active in each round. The auditor does not observe $a_{i,t}$, nor the primitives $\{c_a, q_a\}$; at most, the auditor maintains that actions are ordered as above and that success probabilities are distinct. This partial observability mirrors practical audit settings: posted pay rules and performance metrics are recorded, but the platform's effort technology and workers' disutility are not.

Because α_t is common across agents, the cross-sectional average success rate is a natural sufficient statistic for the state of behavior at time t:

$$\hat{s}_t = \frac{1}{N} \sum_{i=1}^{N} y_{i,t}.$$

When N is large, $\hat{s}_t$ concentrates around the population success probability induced by the (unobserved) action distribution. In this sense, the panel structure supplies a "many small experiments" environment in which time variation in α_t can be traced through to time variation in aggregate success, even though individual actions remain latent.

Mean-based learning and the historical-average approximation. The central behavioral assumption is that agents do not necessarily best respond myopically to α_t each round, but instead adapt using a mean-based no-regret learning algorithm over the finite action set. We impose this in a way that is both flexible (many algorithms qualify) and operational (it yields inequalities with explicit slack). Informally, mean-based learning requires that an action whose cumulative payoff is far behind the best action is played only with small probability. One convenient formulation is: there exists a slack parameter $\gamma(T) = o(1)$ such that, for each agent i and time t, if the cumulative payoff of action a is more than $\gamma(T)T$ below the maximum cumulative payoff across actions, then the probability that the learner chooses a at time t is at most $\gamma(T)$. This property is weaker than specifying a particular update rule, yet strong enough to imply that play concentrates on (approximate) maximizers of average payoffs.

In our setting, expected payoffs have the structure $u^A(\alpha, a) = \alpha q_a - c_a$. A key implication is that, when the learner compares actions using cumulative realized payoffs over time, the incentive component effectively aggregates $\{\alpha_s\}$ through its historical average

$$\bar{\alpha}_t = \frac{1}{t}\sum_{s=1}^{t} \alpha_s.$$

Thus, in large T regimes where the realized averages track their expectations, the mean-based condition implies that the action(s) played with non-negligible probability at time t must be near best responses to $\bar{\alpha}_t$, rather than to the instantaneous α_t. This is the economic channel that makes dynamic contracts potentially manipulative: by shaping $\bar{\alpha}_t$ slowly, the platform can influence behavior even while reducing contemporaneous pay.

Breakpoints and ordering. The strict ordering of $\{q_a\}$ and $\{c_a\}$ implies that indifference between adjacent actions occurs at a unique contract level

$$\alpha_{a-1,a} = \frac{c_a - c_{a-1}}{q_a - q_{a-1}}, \qquad a = 2, \ldots, n.$$

These "breakpoints" summarize the incentive thresholds at which the agent's preferred action changes when evaluated against a scalar incentive index. Although the auditor does not observe $\alpha_{a-1,a}$ directly, the combination of (i) ordered and separated success probabilities and (ii) concentration of play under mean-based learning will allow us to infer when the population is effectively in a regime where one action dominates, and to bound the implied breakpoint locations using only $\{\alpha_t\}$ and $\{\hat{s}_t\}$.

Optional extension: churn. Many platform settings feature turnover: workers enter and exit, and the relevant panel is unbalanced. We therefore allow for an optional churn process in which each agent exits after each round with (observed) hazard $h \in (0, 1)$, independent of current outcomes conditional on the history. The auditor then observes outcomes only for the set of active agents in each round, and $\hat{s}_t$ is computed over those active units. Under independent churn, the main effect is informational: late-round behavior is observed for fewer agents, increasing noise and reducing the visibility of slow-moving learning dynamics. Our core restrictions are robust to this extension, but their statistical power depends on effective sample sizes that decline with h.

The next step is to translate these primitives into restrictions on observable time paths: under mean-based learning, $\bar{\alpha}_t$ becomes the relevant state variable, and changes in $\hat{s}_t$ reveal when the population transitions across incentive regions associated with different actions.

4 Continuous-time restrictions as testable implications

Our audit strategy rests on a simple but powerful implication of mean-based learning in environments with linear incentives: when agents compare actions using cumulative payoffs, the relevant incentive index is not the current contract α_t, but the historical average $\bar{\alpha}_t$. This observation converts a potentially complicated dynamic response into a set of static best-response restrictions evaluated at a slowly moving state variable. We then exploit the geometry induced by ordered actions to translate these restrictions into breakpoint-hitting conditions, which become especially sharp under the "free-fall" contract path.

From mean-based learning to best responses to $\bar{\alpha}_t$. Fix a time t. Under mean-based no-regret learning, any action that is sufficiently far behind the best action in cumulative payoff is played with vanishing probability (up to the slack parameter $\gamma(T)$). Because the agent's expected payoff under action a is $u^A(\alpha, a) = \alpha q_a - c_a$, the payoff differences that govern learning aggregate the incentive terms through $\sum_{s \leq t} \alpha_s$, and hence through $\bar{\alpha}_t$. In a continuous-time proxy (or, equivalently, a large-T regime in which empirical averages are close to expectations), the mean-based condition is well approximated by the requirement that the action(s) played with non-negligible probability at time t lie in the best-response correspondence

$$BR(\bar{\alpha}_t) \in \arg \max_{a \in \{1,\ldots,n\}} \{\bar{\alpha}_t q_a - c_a\}.$$

For auditing, we do not need exact maximization; rather, we need inequalities that tolerate learning slack and finite-sample noise. A convenient reduced-form statement is: there exists an error level $\delta_{N,T}$ such that any action a that has non-negligible population weight at time t must satisfy the pairwise inequalities

$$\bar{\alpha}_t q_a - c_a \geq \bar{\alpha}_t q_b - c_b - \delta_{N,T} \qquad \text{for all } b \in \{1, \ldots, n\}, \tag{1}$$

where $\delta_{N,T} = O(\gamma(T)) + O_P(N^{-1/2})$. The first term reflects that learning algorithms may keep occasionally sampling suboptimal actions; the second term reflects that the auditor sees realized outcomes rather than expectations, so empirical payoffs and empirical success rates concentrate only at rate $1/\sqrt{N}$ each period. In what follows, we treat $\delta_{N,T}$ as an explicit tolerance in moment inequalities.

Breakpoint regions and "boundary switching." The ordered structure of (c_a, q_a) implies a one-dimensional threshold characterization. Indifference between adjacent actions $a-1$ and a occurs at the unique breakpoint

$$\alpha_{a-1,a} = \frac{c_a - c_{a-1}}{q_a - q_{a-1}}, \qquad a = 2, \ldots, n.$$

Because q_a is strictly increasing, the slopes of $\alpha q_a - c_a$ are strictly ordered, and hence the argmax over actions changes only by moving between adjacent actions as α varies. Consequently, in the idealized best-response model, action a is optimal if and only if $\bar{\alpha}_t$ lies in the interval

$$\bar{\alpha}_t \in [\alpha_{a-1,a}, \alpha_{a,a+1}], \qquad (\alpha_{1,2} \text{ finite}, \ \alpha_{n,n+1} = \infty).$$

Under the approximate condition (1), these intervals become "fuzzy" by an amount controlled by $\delta_{N,T}$: if the population concentrates on action a at time t, then $\bar{\alpha}_t$ must lie within a $\delta_{N,T}$-neighborhood of the corresponding breakpoint region. Intuitively, the population cannot persistently exert an action that is sharply dominated at the prevailing historical average incentive, because mean-based learning would drive its probability weight down.

This geometry also yields a sharp prediction about *when* behavior changes. Suppose the population transitions from predominantly playing action a to predominantly playing action $a-1$. In the continuous-time proxy, such a switch can occur only when the relevant state variable $\bar{\alpha}_t$ approaches the boundary at which the two actions are approximately tied, i.e.,

$$\bar{\alpha}_t \approx \alpha_{a-1,a} \qquad \text{at the switching time.}$$

We emphasize that this is a restriction about the *average* contract, not about the current α_t. This is precisely what makes certain dynamic contracts difficult to justify under benign interpretations: the platform may cut current incentives while still keeping $\bar{\alpha}_t$ in a region that sustains high effort for some time.

The free-fall contract and its hyperbolic signature. The free-fall hypothesis posits a particularly stark dynamic: the platform pays a constant success bonus α_0 up to an (unknown) time t_0, then pays zero thereafter,

$$\alpha_t = \alpha_0 \mathbf{1}\{t \leq t_0\}.$$

This path has a distinctive implication for the historical average. For $t \leq t_0$, the average is constant, $\bar{\alpha}_t = \alpha_0$. For $t > t_0$, the average decays deterministically as

$$\bar{\alpha}_t = \frac{1}{t} \sum_{s=1}^{t_0} \alpha_0 = \alpha_0 \frac{t_0}{t}. \tag{2}$$

Thus, after the switch, $\bar{\alpha}_t$ follows a hyperbola: the product $t\bar{\alpha}_t$ is constant and equal to $\alpha_0 t_0$. This hyperbolic decay is the core "signature" of free-fall manipulation, because it pins down the entire path of the incentive state variable using only two scalars (α_0, t_0).

Combining (2) with boundary switching yields a quantitative prediction for the timing of action downgrades. Let t_a denote the (approximate) time at which the population drops from action a to action $a-1$ during the post-t_0 phase. The boundary condition implies $\bar{\alpha}_{t_a} \approx \alpha_{a-1,a}$. Under free-fall, this becomes

$$\alpha_{a-1,a} \approx \alpha_0 \frac{t_0}{t_a} \qquad \Longleftrightarrow \qquad t_a \approx \frac{\alpha_0 t_0}{\alpha_{a-1,a}}.$$

Two testable implications follow. First, switch times are inversely proportional to implied breakpoints: higher breakpoints (harder-to-incentivize actions) must be abandoned sooner as $\bar{\alpha}_t$ decays. Second, across multiple switches, the products $t_a \alpha_{a-1,a}$ must be (approximately) constant:

$$t_a \alpha_{a-1,a} \approx \alpha_0 t_0 \qquad \text{for all observed } a \text{ that are crossed post-} t_0.$$

In other words, once we infer a sequence of action drops from the outcome data, free-fall predicts that the associated boundary hits line up on a single constant $\alpha_0 t_0$. This is a rigid cross-equation restriction: while many contracting policies can generate declining performance, few can generate a *coherent hyperbolic* pattern in the average incentive that matches multiple discrete drops in a way consistent with mean-based learning.

Observable consequences and the role of approximation. The auditor does not observe actions, costs, or the breakpoints $\alpha_{a-1,a}$. Nevertheless, the restrictions above connect observables to latent structure via two channels. The contract path $\{\alpha_t\}$ is observed, so $\bar{\alpha}_t$ is observed mechanically. Outcome data identify when the population success rate changes; under separation of $\{q_a\}$, persistent shifts in $\hat{s}_t$ correspond (in large N) to shifts in the dominant action, at least up to a short transition window during which learners may mix between adjacent actions.

All restrictions are approximate for three reasons that are intrinsic to the environment. First, mean-based learning permits small-probability exploration, captured by $\gamma(T)$. Second, even if all agents played the same action, $\hat{s}_t$ is a noisy estimate of its success probability, with noise of order $1/\sqrt{N}$. Third, the mapping from changes in $\hat{s}_t$ to changes in actions is only sharp when the success probabilities are sufficiently separated (the minimum gap Δ_q governs this). Our approach therefore treats free-fall not as an exact parametric model but as a set of inequalities with slack $\delta_{N,T}$, which we will use to build specification tests and conservative bounds in the next sections.

5 Identification strategy: from outcomes to latent actions and implied breakpoints

The restrictions in Section 4 are expressed in terms of the latent action path and the (also latent) primitives $\{(q_a, c_a)\}_{a=1}^n$. Our goal in this section is to explain how, under separation and a large-population approximation, we can translate observed outcomes and contracts into (i) an estimated segmentation of time into intervals in which the population predominantly plays a single action, and (ii) a corresponding set of implied breakpoint constraints for $\{\alpha_{a-1,a}\}$. Throughout, we emphasize what is point-identified versus what is only set-identified, since the auditor does not observe costs and does not directly observe actions.

Step 1: contracts identify the incentive state $\bar{\alpha}_t$ mechanically. Because the auditor observes the posted sequence $\{\alpha_t\}_{t\le T}$, the historical average

$$\bar{\alpha}_t \;=\; \frac{1}{t}\sum_{s=1}^{t}\alpha_s$$

is directly observed. This is the key state variable in the mean-based restrictions: once we work with $\bar{\alpha}_t$, no further structural assumptions are required to summarize how past incentives enter learning dynamics. In particular, under any candidate model (free-fall or not), the path $\{\bar{\alpha}_t\}$ is known to the auditor, and thus any proposed mapping from incentives to actions must be consistent with $\bar{\alpha}_t$ lying in the appropriate breakpoint region over time.

Step 2: outcomes identify a low-dimensional latent regime process (change-points). Let $\hat{s}_t = \frac{1}{N}\sum_{i=1}^{N} y_{i,t}$ denote the cross-sectional success rate. Conditional on the (unobserved) population distribution over actions at time t, $\hat{s}_t$ concentrates around the mixture success probability

$$s_t \;=\; \sum_{a=1}^{n} p_{a,t} q_a, \qquad p_{a,t} := \Pr(a_{i,t} = a).$$

Mean-based learning implies that, away from indifference boundaries, most mass sits on a best response to $\bar{\alpha}_t$, so that $p_{a,t}$ is near one for a single action a (and near zero for the rest), up to slack $\gamma(T)$. Near a boundary $\alpha_{a-1,a}$, mixing between two adjacent actions may persist for a short window. As a reduced-form implication, $t \mapsto s_t$ is approximately piecewise constant, with jumps when $\bar{\alpha}_t$ crosses a breakpoint region. Therefore, we can treat the time series $\{\hat{s}_t\}$ as a noisy observation of a piecewise-constant signal and estimate change-points $\tau_1 < \cdots < \tau_K$ using standard change-point methods (e.g., penalized least squares, CUSUM, wild binary segmentation), chosen to control false positives under sampling noise of order $N^{-1/2}$.

Operationally, for each candidate segmentation we compute segment means

$$\hat{q}_k \;:=\; \frac{1}{\tau_k - \tau_{k-1}} \sum_{t=\tau_{k-1}+1}^{\tau_k} \hat{s}_t, \qquad k = 1, \ldots, K,$$

with $\tau_0 := 0$. Under large N, $\hat{q}_k$ concentrates around the dominant success probability in that segment, except for boundary windows where it approximates a convex combination of two adjacent q's.

Step 3: separation links segments to an ordered action index (up to relabeling). The separation condition $q_1 < q_2 < \cdots < q_n$ with minimum gap $\Delta_q > 0$ turns the segmentation into information about actions. Intuitively, if two segments have sufficiently different mean success rates, they cannot correspond to the same underlying action. Moreover, because linear contracts preserve a one-dimensional ordering of best responses, the sequence of dominant actions over time must move through adjacent actions as $\bar{\alpha}_t$ drifts. This yields two practical identification consequences.

First, within each segment k, the auditor can assign an *action rank* based on $\hat{q}_k$: higher $\hat{q}_k$ corresponds to higher action index, provided the difference exceeds a tolerance level reflecting sampling error and learning slack. Second, across time, the path of inferred ranks must evolve locally (switching by at most one rank at a time), because best responses change only through adjacent actions under the ordered (q_a, c_a) structure. These two features help distinguish genuine action switches from spurious jumps in $\hat{s}_t$ due to noise, and they discipline the segmentation choice when multiple change-point configurations fit the data similarly well.

We stress a limitation: without observing $\{q_a\}$, the auditor typically cannot *name* the action a in an absolute sense (e.g., whether a given segment corresponds to $a = 3$ or $a = 4$ in the platform's internal taxonomy). What is identified from outcomes alone is an ordered set of latent regimes and their associated success probabilities, up to a monotone relabeling consistent with separation.

Step 4: mapping segments to breakpoint constraints using $\bar{\alpha}_t$. Once we have an estimated segmentation, mean-based optimality converts each segment into restrictions on $\bar{\alpha}_t$. Suppose segment k is mapped to a dominant action a_k (again, an ordered label). Then the breakpoint-region characterization implies that for all t in that segment,

$$\bar{\alpha}_t \in [\alpha_{a_k-1,a_k},\ \alpha_{a_k,a_k+1}]$$

up to a tolerance $\delta_{N,T}$. Since $\bar{\alpha}_t$ is observed, each segment yields an *interval feasibility condition* on the unknown breakpoints: the realized $\bar{\alpha}_t$ values in the segment must fit inside the corresponding breakpoint region. A convenient way to express this is via the segment extrema

$$\underline{\alpha}_k := \min_{t\in(\tau_{k-1},\tau_k]} \bar{\alpha}_t, \qquad \overline{\alpha}_k := \max_{t\in(\tau_{k-1},\tau_k]} \bar{\alpha}_t,$$

which are known. The mean-based restriction then implies (approximately)

$$\alpha_{a_k-1,a_k} \lesssim \overline{\alpha}_k, \qquad \alpha_{a_k,a_k+1} \gtrsim \underline{\alpha}_k,$$

where $\lesssim, \gtrsim$ absorb $\delta_{N,T}$. Thus, even though $\alpha_{a-1,a}$ depends on unobserved costs and success probabilities, the data bound each breakpoint above and below whenever the corresponding action is sustained over a nontrivial time interval. The boundaries are sharpest when $\bar{\alpha}_t$ drifts substantially within a segment: then $[\underline{\alpha}_k, \overline{\alpha}_k]$ is wide, leaving little room for a breakpoint region that could rationalize the segment.

What is point-identified, and what remains set-identified? From the auditor's perspective, $\{\bar{\alpha}_t\}$ is point-identified by construction, and the change-points $\{\tau_k\}$ and segment means $\{\hat{q}_k\}$ are consistently estimable as $N \to \infty$ under standard conditions for piecewise-constant signals. Beyond that, identification is partial.

On the outcome side, the distinct success probabilities $\{q_a\}$ are at best *set-identified* unless the data traverse enough regimes to pin down each q_a directly (and even then, boundary mixing can blur exact equality). What we can recover robustly is an ordered list of distinct regime means appearing in the sample, together with confidence bands whose width scales as $O_P(N^{-1/2})$.

On the incentive side, the breakpoints $\{\alpha_{a-1,a}\}$ are also generally *set-identified*. Each observed segment provides inequalities that bound adjacent breakpoints, but without observing costs we cannot invert $\alpha_{a-1,a} = (c_a - c_{a-1})/(q_a - q_{a-1})$ to recover $\{c_a\}$, nor can we determine breakpoints corresponding to actions never played. Point identification of a given breakpoint requires that the data include a sufficiently sharp switch where $\bar{\alpha}_t$ crosses the boundary and the mixing window is negligible relative to $\delta_{N,T}$; otherwise, the switch only pins the breakpoint to an interval whose width reflects learning slack, sampling noise, and the speed of drift in $\bar{\alpha}_t$.

These identification statements are exactly what we need for auditing: the subsequent specification tests will not rely on point estimates of $\{q_a, c_a\}$. Instead, they exploit the fact that any mean-based explanation must satisfy the breakpoint-region inequalities for $\bar{\alpha}_t$ *and*, under free-fall, must additionally satisfy the hyperbolic timing structure linking multiple switches.

6 Specification tests: mean-based compatibility and free-fall structure

Having translated the data into an estimated regime process (change-points and segment means) and a collection of breakpoint-region constraints indexed by those regimes, we now describe two related specification tests. The first asks whether the joint path $\{\alpha_t, \hat{s}_t\}_{t\leq T}$ is compatible with *some* mean-based principal–agent model with ordered actions (allowing an arbitrary contract sequence). The second is sharper: it asks whether the same data are compatible with the *one-switch free-fall* manipulation pattern (an initial constant contract followed by zero), which imposes additional cross-segment restrictions on the *timing* of action drops.

6.1. A moment-inequality test for mean-based compatibility. The mean-based restrictions are inherently inequality-based because (i) learners may mix or explore with probability at most $\gamma(T)$, and (ii) the auditor infers actions only through noisy outcome averages. Accordingly, our null is a feasibility statement: there exist latent primitives and a latent action labeling under which the observed data satisfy the breakpoint-region implications up to slack.

Fix a candidate segmentation $\hat{\tau}_1 < \cdots < \hat{\tau}_{\hat{K}}$ and the induced set of segments $k = 1, \ldots, \hat{K}$. For each segment define $\underline{\alpha}_k$ and $\overline{\alpha}_k$ as the within-segment extrema of $\bar{\alpha}_t$ (computed mechanically from $\{\alpha_t\}$). Let $\hat{q}_k$ be the segment mean of $\hat{s}_t$. Separation implies that the segments can be assigned an ordered action rank; we operationalize this by requiring that there exist values $q_1 < \cdots < q_m$ (where $m \leq n$ is the number of distinct success rates visited in the sample) such that each $\hat{q}_k$ lies near one of these values, with tolerance b_N of order $N^{-1/2}$. We then impose the mean-based best-response inequalities in a way that does not require observing costs: for each segment k, there must exist adjacent breakpoints $\alpha_{j-1,j} \le \alpha_{j,j+1}$ for the associated rank j such that

$$\alpha_{j-1,j} \;\le\; \overline{\alpha}_k + \delta_{N,T}, \qquad \alpha_{j,j+1} \;\ge\; \underline{\alpha}_k - \delta_{N,T}, \tag{3}$$

where $\delta_{N,T}$ absorbs both learning slack $O(\gamma(T))$ and outcome noise $O_P(N^{-1/2})$. Intuitively, (3) says that the realized $\bar{\alpha}_t$ values in a segment must fit inside some breakpoint region consistent with an approximately optimal action.

This yields a finite collection of inequalities of the generic form $g_\ell(\theta) \le 0$, where θ collects nuisance objects (the segment-level $\{q_j\}$, the breakpoints $\{\alpha_{j-1,j}\}$, and the segment-to-rank assignment). The mean-based compatibility null can therefore be written as

$$H_0^{\mathrm{MB}} : \; \exists\, \theta \in \Theta \text{ s.t. } \max_\ell g_\ell(\theta) \le 0.$$

We reject H_0^{MB} when the inequalities are jointly infeasible beyond the allowed slack. In practice, we implement this as a minimization of the worst violation,

$$\hat{V}_{\mathrm{MB}} \;:=\; \inf_{\theta \in \Theta} \; \max_\ell \big[g_\ell(\theta)\big]_+, \qquad [x]_+ := \max\{x, 0\},$$

and compare $\hat{V}_{\mathrm{MB}}$ to a critical value that accounts for sampling variation in $\{\hat{q}_k\}$ and in the estimated change-points. Two remarks are important for auditing. First, $\hat{V}_{\mathrm{MB}}$ is conservative with respect to unknown costs because costs enter only through the breakpoints, which we treat as free (subject to ordering) within Θ. Second, failure to reject H_0^{MB} should be read narrowly: it means the data can be rationalized by *some* ordered-action, mean-based model, not that any particular platform behavior is validated.

6.2. A sharper test for one-switch free-fall. Mean-based compatibility alone does not distinguish benign contracting from free-fall manipulation. The free-fall hypothesis adds a rigid structure: after an (unknown) time t_0, the posted contract is zero and thus $\bar{\alpha}_t = \alpha_0 t_0 / t$ declines hyperbolically. Combined with the breakpoint-hitting logic, this implies that when the population drops from one action to the next, the historical average at the change-point must approximately equal the corresponding breakpoint, and hence the products "time × historical average" must line up across multiple drops.

Let $\hat{\tau}_1 < \cdots < \hat{\tau}_{\hat{K}}$ denote estimated change-points in $\hat{s}_t$, and focus on those $\hat{\tau}_k$ at which the inferred regime mean decreases (interpreted as an action drop). For each such k, define the statistic

$$\hat{z}_k \;:=\; \hat{\tau}_k \, \bar{\alpha}_{\hat{\tau}_k}.$$

Under the free-fall path, if $\hat{\tau}_k > t_0$ and $\hat{\tau}_k$ corresponds to a boundary crossing, then $\bar{\alpha}_{\hat{\tau}_k} \approx \alpha_{a-1,a}$ and simultaneously $\bar{\alpha}_{\hat{\tau}_k} = \alpha_0 t_0 / \hat{\tau}_k$, so

$$\hat{z}_k \;\approx\; \alpha_0 t_0, \tag{4}$$

up to the same tolerance $\delta_{N,T}$ plus change-point localization error. Condition (4) is attractive because it avoids explicit estimation of $\alpha_{a-1,a}$: the breakpoints drop out, leaving a common constant across drops.

We therefore define a free-fall discrepancy

$$\hat{V}_{\mathrm{FF}} \;:=\; \min_{c \in [0,T]} \; \max_{k \in \mathcal{D}} \left| \hat{z}_k - c \right|,$$

where $\mathcal{D}$ indexes the set of detected decreases in regime means. Small $\hat{V}_{\mathrm{FF}}$ indicates that the observed action drops occur at times consistent with a single hyperbolic regime governed by a common $\alpha_0 t_0$. To incorporate the defining *one-switch* feature, we additionally require that there exists a t_0 such that α_t is approximately constant for $t \le t_0$ and approximately zero for $t > t_0$ (allowing a tolerance band to accommodate discretization or small implementation noise). This second check uses the observed $\{\alpha_t\}$ directly; the role of $\hat{V}_{\mathrm{FF}}$ is to ensure that, even if $\{\alpha_t\}$ is only approximately free-fall, the *outcome-implied* switch times are consistent with the implied $\bar{\alpha}_t$ geometry.

6.3. Size, power, and what drives detectability. Under benign regimes such as static contracting, $\bar{\alpha}_t$ is approximately constant and hence $\hat{z}_k$ grows linearly with $\hat{\tau}_k$ rather than remaining constant; thus multiple action drops generically violate (4). This is the core source of power in the free-fall test: it leverages cross-drop restrictions rather than within-segment fit. Power increases with N (sharper $\hat{s}_t$ and more accurate change-point detection) and with Δ_q (cleaner separation of regimes). Conversely, larger learning slack $\gamma(T)$ weakens the mapping from $\bar{\alpha}_t$ to a single dominant action, inflating $\delta_{N,T}$ and making both tests more permissive.

For size control, we treat both tests as moment-inequality problems with nuisance parameters and rely on calibration that is robust to the weak dependence induced by learning. A practical approach is to (i) fix the estimated segmentation procedure, (ii) compute self-normalized standard errors for segment means using the cross-section (variance $\hat{q}_k(1-\hat{q}_k)/N$ within each t, aggregated over t in the segment), and (iii) use a block bootstrap over time for the statistics built from $\{\hat{\tau}_k\}$ and $\{\bar{\alpha}_t\}$, which captures the additional randomness from change-point localization. We then select critical values in the style of generalized moment selection (e.g., focusing on inequalities close to binding) to avoid excessive conservatism when many constraints are slack.

Finally, finite-sample implementation requires choosing the tolerance $\delta_{N,T}$. We recommend an explicit decomposition $\delta_{N,T} = \delta_N + \delta_T$, where δ_N is set from binomial concentration (e.g., a normal or Hoeffding bound scaled by $N^{-1/2}$ and the segment length) and δ_T is set to a user-chosen upper bound on $\gamma(T)$ reflecting the assumed learning algorithm class. This makes the auditing posture transparent: tighter assumptions about learner slack translate mechanically into a more demanding specification test.

7 Estimating manipulation magnitude: breakpoints and surplus transfer

The specification tests in Section 6 are deliberately formulated as *feasibility* checks. When the data are compatible with mean-based learning and, in particular, when we do not reject one-switch free-fall, the natural next question for an auditor is quantitative: *how large is the implied manipulation?* In our setting this is not a structural exercise in recovering costs; rather, we use the model-implied geometry to (i) reconstruct an *implied breakpoint sequence* (hence an incentive–outcome response map) and (ii) compute a *conservative lower bound* on the platform's incremental payoff relative to the best static linear contract. We also describe uncertainty quantification that is transparent about what is and is not identified from $\{\alpha_t, y_{i,t}\}$.

7.1. From change-points to an implied breakpoint sequence. Under separated success probabilities, a decrease in the segment mean of $\hat{s}_t$ is naturally interpreted as a downward move in the dominant action. Under mean-based learning, such a move should occur when the historical average $\bar{\alpha}_t$ crosses the relevant indifference boundary. Accordingly, if $\hat{\tau}_k$ is an estimated time at which the regime mean decreases, we associate the *implied breakpoint hit* with the observable quantity $\bar{\alpha}_{\hat{\tau}_k}$.

To make this operational, let $\mathcal{D} \subseteq \{1, \ldots, \hat{K}\}$ index the detected decreases in regime means. For each $k \in \mathcal{D}$, define

$$\hat{b}_k := \bar{\alpha}_{\hat{\tau}_k}, \qquad \hat{z}_k := \hat{\tau}_k \, \bar{\alpha}_{\hat{\tau}_k}.$$

Under an approximately free-fall path and approximate best-response play, we expect $\hat{b}_k$ to be close to a true breakpoint $\alpha_{a-1,a}$, and $\hat{z}_k$ to be close to the constant $\alpha_0 t_0$ whenever $\hat{\tau}_k > t_0$. In practice, multiple decreases can occur after t_0, so we summarize $\{\hat{z}_k\}_{k \in \mathcal{D}}$ by a robust location estimate,

$$\hat{c} \in \arg \min_{c \in [0,T]} \max_{k \in \mathcal{D}} |\hat{z}_k - c|,$$

which matches the free-fall discrepancy already used for testing. Conditional on not rejecting free-fall, $\hat{c}$ provides a one-dimensional summary of the *speed* of the hyperbolic decline $\bar{\alpha}_t \approx \hat{c}/t$ in the post-switch period.

We then assemble an ordered implied breakpoint sequence by sorting $\{\hat{b}_k\}_{k \in \mathcal{D}}$ from largest to smallest. When the data exhibit clear step-downs in $\hat{s}_t$, these sorted values correspond to successive $\alpha_{a-1,a}$'s visited along the trajectory. When some action levels are skipped (e.g., agents jump from a to $a-2$ due to coarse time resolution or imperfect change-point localization), the sorted $\hat{b}_k$'s should be interpreted as a *subsequence* of the full breakpoint ladder. This partial identification is still sufficient for our payoff bound below, because the static benchmark depends only on *visited* actions and their implied minimum incentives.

7.2. Reconstructing the static benchmark without costs. A key advantage of linear contracts with ordered actions is that the principal's expected payoff at action a is

$$u^P(\alpha, a) = (1-\alpha)q_a,$$

and if a given α induces action a, then any larger α in the same best-response region leaves q_a unchanged while reducing $(1-\alpha)$. Hence, among contracts that induce action a, the best *static* contract is (approximately) the smallest α that still induces a, namely the breakpoint $\alpha_{a-1,a}$. This implies the benchmark

$$U^P_{\text{static}} \;=\; T \cdot \max_{a\geq 1} \,(1-\alpha_{a-1,a})q_a, \qquad (\alpha_{0,1} := 0), \tag{5}$$

where the maximization is over the set of actions that are feasible in the model (and, empirically, plausibly visited).

Equation (5) is useful for auditing because it eliminates costs entirely: all cost information is subsumed by the breakpoints, which are in turn tied to observable $\bar{\alpha}_t$ at action drops. Concretely, for each visited post-switch drop $k \in \mathcal{D}$ we treat $\hat{b}_k = \bar{\alpha}_{\hat{\tau}_k}$ as an estimate of some $\alpha_{a-1,a}$, and we treat the corresponding segment mean $\hat{q}_k$ (or the pre/post segment means around $\hat{\tau}_k$) as estimates of the relevant q-levels. This yields an *empirical envelope* over candidate action levels:

$$\widehat{U}^P_{\text{static}} \;:=\; T \cdot \max_{j\in\widehat{\mathcal{J}}} \,(1-\hat{\alpha}^{\downarrow}_j)\hat{q}^{\uparrow}_j,$$

where $\widehat{\mathcal{J}}$ indexes distinct outcome levels inferred from the segmentation, $\hat{\alpha}^{\downarrow}_j$ is the inferred "minimum incentive" for level j (constructed from the relevant $\hat{b}_k$'s, with monotonicity enforced by isotonic regression if needed), and $\hat{q}^{\uparrow}_j$ is the inferred success probability at that level.

The dynamic payoff is directly observed from outcomes and posted contracts:

$$\widehat{U}^P_{\text{dyn}} \;:=\; \sum_{t=1}^{T}(1-\alpha_t)\hat{s}_t, \qquad \hat{s}_t = \frac{1}{N}\sum_{i=1}^{N} y_{i,t},$$

which is a per-agent payoff; multiplying by N gives the platform aggregate. We then report the implied incremental gain

$$\widehat{\Delta}^P \;:=\; \widehat{U}^P_{\mathrm{dyn}} - \widehat{U}^P_{\mathrm{static}}. \tag{6}$$

To maintain a conservative auditing posture, we interpret $\widehat{U}^P_{\mathrm{static}}$ as an *upper* bound on what a well-chosen static contract could have achieved, so $\widehat{\Delta}^P$ is a *lower* bound on the incremental payoff attributable to the dynamic pattern, subject to the maintained mean-based model.

7.3. Uncertainty quantification and conservative bands. Two sources of uncertainty matter: (i) sampling noise in $\hat{s}_t$ (and hence in segment means $\hat{q}_k$), and (ii) error in locating change-points $\hat{\tau}_k$, which propagates into $\hat{b}_k = \bar{\alpha}_{\hat{\tau}_k}$ and $\hat{z}_k$. Both are compounded by learning slack $\gamma(T)$, which we treat as an explicit tolerance component rather than something to be "estimated away."

For outcome noise, within a segment k of length L_k, a convenient approximation is

$$\hat{q}_k \approx \mathcal{N}\left(q_k,\ \frac{q_k(1-q_k)}{NL_k}\right),$$

which yields Wald-type intervals; for a more conservative bound, we can use a Hoeffding inequality uniformly over t and aggregate within segments. For breakpoint uncertainty, we recommend a block bootstrap over time applied to the sequence $\{\hat{s}_t\}_{t\le T}$ (keeping $\{\alpha_t\}$ fixed), re-running the segmentation and recomputing $\{\hat{\tau}_k\}$, $\{\hat{b}_k\}$, and $\widehat{\Delta}^P$ in each bootstrap draw. This captures both change-point localization variability and weak time dependence induced by learning.

To produce a *conservative* lower confidence bound for Δ^P, we exploit monotonicity: $(1-\alpha)q$ is increasing in q and decreasing in α. Thus an upper bound on the static benchmark uses an *upper* confidence band for q and a *lower* confidence band for the inducing breakpoint. Let $[\hat{q}^L_j, \hat{q}^U_j]$ and $[\hat{\alpha}^L_j, \hat{\alpha}^U_j]$ denote confidence intervals widened by the slack $\delta_{N,T}$ used in the specification stage. Define

$$\widehat{U}^{P,U}_{\mathrm{static}} \;:=\; T\cdot \max_{j\in\widehat{\mathcal{J}}} (1-\hat{\alpha}^L_j)\hat{q}^U_j, \qquad \widehat{U}^{P,L}_{\mathrm{dyn}} \;:=\; \sum_{t=1}^{T}(1-\alpha_t)\hat{s}^L_t,$$

with $\hat{s}^L_t$ a pointwise lower confidence bound for s_t. Then

$$\widehat{\Delta}^{P,L} \;:=\; \widehat{U}^{P,L}_{\mathrm{dyn}} - \widehat{U}^{P,U}_{\mathrm{static}}$$

is a conservative lower bound on incremental payoff that is explicit about sampling error and about the assumed learning slack. When $\widehat{\Delta}^{P,L}$ is economically material (e.g., large relative to baseline margins or policy thresholds), the auditor can report not only that the free-fall pattern is statistically compatible, but also that it plausibly generated nontrivial surplus reallocation.

These estimation steps are straightforward to stress-test in controlled environments. We therefore turn next to simulations that generate synthetic data from mean-based learners under free-fall and explore robustness to exploration, noise, heterogeneity, and partial observability.

8 Simulation study: recovering free-fall signatures in synthetic data

Our theoretical restrictions are deliberately austere: the auditor sees posted contracts and realized outcomes, while actions and costs remain latent. A simulation study is therefore useful for two reasons. First, it verifies that the specification tests in Section 6 behave as advertised in finite samples (size control under benign contracting, power under free-fall-compatible manipulation). Second, it clarifies how the economic primitives that are unobserved in practice—exploration in learning, heterogeneity, and partial observability—map into the tolerances $\delta_{N,T}$ and into the precision of the implied breakpoints and surplus-transfer bounds in Section 7.

8.1. Data-generating process (DGP). We generate synthetic panels $\{(\alpha_t, y_{i,t})\}_{i\le N, t\le T}$ from the model primitives. Fix an action set $\{1,\dots,n\}$ with strictly increasing success probabilities $0 = q_1 < q_2 < \cdots < q_n < 1$ and strictly increasing costs $0 = c_1 < c_2 < \cdots < c_n$. The principal posts either (i) the free-fall sequence $\alpha_t = \alpha_0 \mathbf{1}\{t \le t_0\}$ or (ii) a null sequence meant to mimic "benign" contracting (static $\alpha_t \equiv \alpha^*$, piecewise-constant incentives without a hyperbolic-average regime, or smoothly varying incentives with no mass at $\alpha_t = 0$). Agents are i.i.d. and, conditional on action choice $a_{i,t}$, outcomes satisfy

$$y_{i,t} \sim \text{Bernoulli}(q_{a_{i,t}}), \qquad w_{i,t} = \alpha_t y_{i,t}, \qquad \pi^A_{i,t} = w_{i,t} - c_{a_{i,t}}.$$

The auditor constructs $\hat{s}_t = \frac{1}{N}\sum_{i=1}^N y_{i,t}$ and $\bar{\alpha}_t = \frac{1}{t}\sum_{s=1}^t \alpha_s$ exactly as in the empirical procedure.

8.2. Mean-based learners: an implementable proxy. To operationalize mean-based no-regret, we use a learning rule that makes action probabilities a monotone function of empirical average payoffs, with an explicit exploration component. One convenient choice is a multiplicative-weights (Hedge) update with forced exploration: each agent maintains scores $S_{i,t}(a)$ for actions $a \in \{1,\dots,n\}$, updates the chosen action using realized payoff $\pi^A_{i,t}$, and samples at time $t+1$ from

$$p_{i,t+1}(a) \;\propto\; \exp\big(\eta\, S_{i,t}(a)\big), \qquad p_{i,t+1} \leftarrow (1-\varepsilon_t)p_{i,t+1} + \varepsilon_t \cdot \text{Unif}(\{1,\dots,n\}).$$

Here $\eta > 0$ is a learning rate and ε_t governs exploration. In large T, the forced exploration implies a bound of the form $\gamma(T) \asymp \max_t \varepsilon_t$ on the probability assigned to actions that are far behind in average payoff; by varying ε_t we directly stress-test the role of $\gamma(T)$ in our moment inequalities. Because agents observe their own outcomes and payments, the update uses realized utilities, but the auditor does not need to observe any of these internal objects.

8.3. What we measure. Each Monte Carlo replication produces (i) estimated change-points $\{\hat{\tau}_k\}$ from the time series $\{\hat{s}_t\}$ using the same segmentation routine employed in the audit, (ii) the implied breakpoint hits $\hat{b}_k = \bar{\alpha}_{\hat{\tau}_k}$ and the free-fall invariants $\hat{z}_k = \hat{\tau}_k \bar{\alpha}_{\hat{\tau}_k}$ for detected decreases, and (iii) the test outcomes for the free-fall specification checks. We report:

1. *Size*: under null contract paths, the empirical rejection probability at nominal level η.
2. *Power*: under free-fall, rejection probability as a function of N, T, Δ_q, and exploration.
3. *Localization error*: $|\hat{\tau}_k - \tau_k|$ and its impact on $|\hat{b}_k - \alpha_{a-1,a}|$.
4. *Magnitude recovery*: bias and dispersion of $\widehat{\Delta}^P$ and of the conservative lower bound $\widehat{\Delta}^{P,L}$ from Section 7.

These objects align with the auditor's questions: (a) do we falsely flag benign contracting, (b) do we detect manipulation when present, and (c) if compatible, can we quantify a nontrivial payoff gap.

8.4. Baseline calibration and expected patterns. In a baseline design we choose moderate n (e.g., $n \in \{3, 5\}$), separated q-levels (e.g., evenly spaced so Δ_q is nontrivial), and costs chosen so that multiple breakpoints $\alpha_{a-1,a} \in (0, 1)$ exist. We set (α_0, t_0) so that $\bar{\alpha}_t$ crosses at least two breakpoints post-switch. The qualitative predictions are immediate: holding all else fixed, larger N tightens concentration of $\hat{s}_t$ around its mean at rate $1/\sqrt{N}$, improving both change-point detection and the mapping from $\bar{\alpha}_{\hat{\tau}_k}$ to breakpoints; larger T increases the number of boundary crossings and sharpens the hyperbolic pattern in $\bar{\alpha}_t$; smaller $\gamma(T)$ makes mixing negligible so the population more cleanly concentrates on adjacent actions.

8.5. Robustness dimensions. We extend the baseline along four dimensions that correspond to the main practical objections an auditor might raise.

Exploration and slow learning. We increase ε_t (constant or slowly decaying) to generate persistent mixing across actions. This tests the necessity of allowing slack $\delta_{N,T} = O(\gamma(T)) + O_P(1/\sqrt{N})$: as $\gamma(T)$ rises, apparent change-points blur, adjacent segments become harder to distinguish when Δ_q is modest, and the free-fall equalities in $\hat{z}_k$ become noisier. The test should maintain size if the slack is widened accordingly, but power falls mechanically.

Outcome noise and common shocks. Beyond Bernoulli variance, we add time shocks ξ_t that shift all success probabilities multiplicatively (e.g., $q_a(t) = \text{logit}^{-1}(\text{logit}(q_a) + \xi_t)$) to capture demand or environment volatility. This probes whether segmentation mistakes are driven by true action shifts or by exogenous shocks. A practical lesson we expect to emerge is that auditors should either control for observable covariates correlated with ξ_t, or use conservative penalties in change-point detection to avoid over-interpreting transient fluctuations as effort drops.

Heterogeneity. We allow agent types m with type-specific costs $c_a^{(m)}$ (and optionally type-specific $q_a^{(m)}$), drawing types i.i.d. at entry. This generates simultaneous mixing across best responses even when each type is internally concentrated. In such environments, $\hat{s}_t$ reflects a mixture of actions across types, so the "step-down" pattern may be attenuated; however, under ordered actions and common $\bar{\alpha}_t$, the aggregate success rate remains monotone in $\bar{\alpha}_t$, so the free-fall invariants can still be informative, albeit with fewer sharp change-points. We also incorporate churn by letting each agent exit with hazard h, replacing them with new entrants whose learning restarts; this truncates late-stage evidence and is predicted to reduce power primarily for later breakpoint hits.

Partial observability and aggregation. Finally, we coarsen the auditor's data to settings common in practice: (i) only aggregate success counts per period are available (no individual panel), (ii) outcomes are observed at weekly rather than daily frequency, and (iii) contracts are observed with rounding or delay. These variants primarily affect the precision of $\hat{s}_t$ and the localization of $\hat{\tau}_k$. The implication is not that the audit fails, but that conservative tolerances and coarser segmentation (fewer allowable change-points) become necessary to preserve size.

8.6. Outputs and interpretation. We present results as heatmaps of rejection probabilities and as event-study plots of $\hat{s}_t$ against $\bar{\alpha}_t$, emphasizing the geometry that motivates the test: under free-fall, decreases in $\hat{s}_t$ align with nearly constant $\hat{z}_k = \hat{\tau}_k \bar{\alpha}_{\hat{\tau}_k}$, whereas under static or smoothly varying contracts they do not. For magnitude, we compare $\widehat{\Delta}^P$ and $\widehat{\Delta}^{P,L}$ to the true simulated payoff difference relative to the best static linear contract. The central practical message is that compatibility tests and conservative bounds can be simultaneously informative when N is large enough to stabilize $\hat{s}_t$ and when learning is not so exploratory that $\gamma(T)$ dominates sampling noise.

The next step is to translate these procedures into an implementable empirical plan: what data an auditor would need, which proxies suffice when full panels are unavailable, and how to operationalize the workflow as a reproducible audit protocol.

9 Empirical application plan: data requirements and an implementable audit protocol

Our theory is intentionally designed for environments in which the auditor does *not* observe effort, costs, or the platform's internal optimization. The empirical objective is therefore modest but operational: using only posted incentives and realized outcomes, we ask whether the joint time series is compatible with (approximately) free-fall dynamic contracting under mean-based learning, and, if compatible, we bound the implied incremental surplus captured by the principal relative to the best static linear benchmark.

9.1. Minimal data and preferred data. At the minimum, the auditor needs a time series of (i) the posted linear incentive $\alpha_t \in [0, 1]$ (or a verifiable proxy for the *marginal* bonus per success), and (ii) realized binary outcomes $y_{i,t}$ or their period aggregates. When individual panels are available, we prefer the full panel $\{y_{i,t}, w_{i,t}\}_{i \le N, t \le T}$ along with a definition of the active set each period (entry/exit timestamps). When only aggregates are available, it suffices to observe success counts $Y_t = \sum_{i=1}^{N_t} y_{i,t}$ and denominators N_t (exposure), so that $\hat{s}_t = Y_t/N_t$ is well-defined. In practice α_t may be multi-dimensional; our recommended mapping is to construct the *effective* marginal incentive for the audited action-outcome pair (e.g., bonus-per-delivery, bonus-per-accepted-job), and to explicitly document any transformations from the raw contract.

9.2. Plausible proxies when contracts are proprietary. If the platform does not disclose α_t directly, we can often recover it from observed payments and outcomes because $w_{i,t} = \alpha_t y_{i,t}$ implies $\alpha_t = \mathbb{E}[w_{i,t} \mid y_{i,t} = 1]$ under correct measurement of the relevant success event. Operationally, we estimate $\hat{\alpha}_t$ as the average bonus paid conditional on success, using all successful transactions in period t. This requires transaction-level payments (or at least mean bonus among successes). If only total payments are observed, the auditor can bound α_t using $\sum_i w_{i,t} \le \alpha_t \sum_i y_{i,t}$ plus institutional constraints (e.g., published maximum bonus). We emphasize that measurement error in α_t enters our restrictions directly through $\bar{\alpha}_t$, so the audit should report sensitivity to plausible rounding and reporting lags.

9.3. Pre-processing: constructing $\hat{s}_t$ and $\bar{\alpha}_t$. Given α_t (or $\hat{\alpha}_t$) and outcomes, we form

$$\hat{s}_t = \frac{1}{N_t}\sum_{i=1}^{N_t} y_{i,t}, \qquad \bar{\alpha}_t = \frac{1}{t}\sum_{s=1}^{t} \alpha_s,$$

with N_t the number of active agents (or exposures) at t. If exposures vary materially over time, we recommend two parallel series: the unweighted $\hat{s}_t$ above, and an exposure-weighted variant over a coarser window (e.g., weekly), which stabilizes variance when N_t is small in some periods. The auditor should also pre-register an exclusion rule for periods with abnormal measurement (platform outages, missing outcome logs) to avoid spurious change-points.

9.4. Step 1: detecting outcome regimes and candidate effort drops. The empirical analog of latent action switching is a change-point structure in $\hat{s}_t$. We implement segmentation on $\{\hat{s}_t\}$ to obtain $\hat{\tau}_1 < \cdots < \hat{\tau}_{\hat{K}}$. Any standard method with a transparent penalty for over-fitting is acceptable (binary segmentation with information criteria, fused lasso, or likelihood-based multiple change-point estimation), provided the auditor reports robustness to the penalty choice. We then classify candidate *drops* as indices k such that $\hat{s}_{\hat{\tau}_k+1} - \hat{s}_{\hat{\tau}_k} < 0$ (after smoothing within segments). Intuitively, these are the moments at which the population appears to switch to a lower-success action.

9.5. Step 2: free-fall invariants and moment-inequality checks. Under free-fall, $\bar{\alpha}_t$ follows a hyperbola after the switch, so breakpoint hits imply near-constancy of the products $z_k = \tau_k \bar{\alpha}_{\tau_k}$ across successive drops. Accordingly, for each detected drop at $\hat{\tau}_k$ we compute

$$\hat{b}_k = \bar{\alpha}_{\hat{\tau}_k}, \qquad \hat{z}_k = \hat{\tau}_k \, \bar{\alpha}_{\hat{\tau}_k}.$$

A simple specification diagnostic is whether $\hat{z}_k$ is approximately constant over multiple drops, up to tolerances that reflect sampling noise and learning slack. The formal procedure uses the moment inequalities implied by best responses to $\bar{\alpha}_t$: within each estimated segment, $\bar{\alpha}_t$ should lie in an interval consistent with a single action (or two adjacent actions) that rationalizes the segment mean $\hat{s}_t$. Because the auditor does not know $\{q_a\}$, we treat them as nuisance parameters constrained only by ordering and separation; inference is therefore set-identified. Practically, we recommend reporting two versions of the test: (i) a conservative version that assumes only a minimal separation Δ_q and allows larger slack $\delta_{N,T}$, and (ii) an informative version that calibrates $\delta_{N,T}$ using placebo periods and the estimated variance of $\hat{s}_t$.

9.6. Step 3: estimating (t_0, α_0) and reconstructing breakpoints. If the invariants support a free-fall interpretation, we estimate the switch time t_0 by searching over candidate t for which α_t becomes (approximately) zero and $\bar{\alpha}_t$ begins a $1/t$ decline. When α_t is noisy, we instead estimate t_0 as the point that best linearizes $t \mapsto t\bar{\alpha}_t$ after t_0. Given $\hat{t}_0$, we estimate $\hat{\alpha}_0$ as the mean of α_t over $t \leq \hat{t}_0$. We then map each drop time $\hat{\tau}_k > \hat{t}_0$ to an implied breakpoint estimate $\widehat{\alpha}_k = \bar{\alpha}_{\hat{\tau}_k}$. Under the model, $\widehat{\alpha}_k$ should correspond to some $\alpha_{a-1,a}$; we do not require labeling the action index a to conduct the core manipulation test, but labeling becomes useful for magnitude calculations.

9.7. Step 4: magnitude and conservative surplus-transfer bounds. To quantify economic significance, we compute the principal's realized payoff using observables:

$$\widehat{U}^P_{\text{dyn}} = \sum_{t=1}^{T} (1 - \alpha_t)\, \hat{s}_t,$$

interpreting $\hat{s}_t$ as the per-agent success rate (or using Y_t directly in totals). To form the best static benchmark, we estimate an empirical response curve $\hat{q}(\alpha)$ by associating each segment with its mean success rate and its corresponding $\bar{\alpha}_t$ range, yielding a piecewise-constant (or monotone-smoothed) mapping. The auditor then computes

$$\widehat{U}^P_{\text{static}} = \max_{\alpha \in [0,1]} \sum_{t=1}^{T} (1 - \alpha)\, \hat{q}(\alpha),$$

and reports $\widehat{\Delta}^P = \widehat{U}^P_{\text{dyn}} - \widehat{U}^P_{\text{static}}$. A conservative lower bound $\widehat{\Delta}^{P,L}$ is obtained by minimizing over all response curves consistent with the moment inequalities (ordering and separation), which we implement by solving a finite-dimensional program over segment-level success rates.

9.8. Reporting, robustness, and limitations. We recommend that an audit report include: (i) plots of $\hat{s}_t$ and $\bar{\alpha}_t$ with estimated change-points; (ii) the sequence $\{\hat{z}_k\}$ with confidence bands obtained by block bootstrap over t; (iii) rejection decisions for both conservative and informative specifications; and (iv) $\widehat{\Delta}^P$ and $\widehat{\Delta}^{P,L}$ with sensitivity to Δ_q, to alternative segmentation penalties, and to controls for observable shocks (seasonality, policy changes). The core limitation is interpretability under rich nonstationarities: if exogenous shocks shift success probabilities directly, change-points in $\hat{s}_t$ may not correspond to effort changes. Our protocol therefore treats covariate adjustment and conservative segmentation as first-order safeguards. Within these constraints, the procedure yields a reproducible workflow that can be executed with either full microdata or aggregates, producing a transparent compatibility assessment and an economically interpretable bound on the potential gains from dynamic manipulation.

10 Discussion and extensions

Our framework isolates a narrow but practically salient channel: with linear pay-for-success incentives and mean-based learning, a platform can induce predictable "effort regimes" that are detectable in aggregate outcome data, even when effort and costs are unobserved. In this section we discuss how far these ideas can be pushed in richer environments that auditors actually face, and what kinds of regulatory questions our restrictions can (and cannot) help answer.

Multi-KPI and multi-dimensional contracts. Many platforms do not pay on a single binary outcome but on a vector of key performance indicators (KPIs): acceptance rates, completion rates, ratings, timeliness, and occasionally penalties (e.g., cancellations). A direct generalization is to treat the agent as choosing an action (or effort allocation) that induces a vector of success probabilities $q_a \in [0,1]^m$, and the platform as posting a linear contract with weights $\boldsymbol{\alpha}_t \in [0,1]^m$, so that per-round utility becomes $u^A(\boldsymbol{\alpha}_t, a) = \boldsymbol{\alpha}_t \cdot q_a - c_a$. Under the same "best response to historical averages" logic, the relevant sufficient statistic is the historical average weight vector $\bar{\boldsymbol{\alpha}}_t$, and action switches occur when $\bar{\boldsymbol{\alpha}}_t$ crosses a supporting hyperplane between two actions. Relative to the scalar case, the geometry is more complex: indifference "breakpoints" become facets of a polyhedral complex, and the auditor should not expect a single hyperbolic invariant like $t\bar{\alpha}_t$ to summarize the dynamics.

That said, much of the operational content survives if the auditor can reduce the multidimensional contract to a one-dimensional *effective* incentive for the KPI that is being audited. A sufficient condition is that (i) the platform varies primarily the weight on one KPI while holding the others fixed (or changes them only slowly), and (ii) the action ordering is essentially one-dimensional in the sense that higher actions shift the audited KPI monotonically and dominate lower actions for sufficiently large marginal incentives on that KPI. In that case, our scalar restrictions can be applied to the implied time series α_t^{eff} for the audited KPI, interpreted as a projection of $\boldsymbol{\alpha}_t$. A complementary robustness check is to repeat the procedure KPI-by-KPI: free-fall manipulation targeted to a headline metric should generate stronger breakpoint-hitting structure for that metric than for auxiliary metrics. Conversely, if effort substitution is central (e.g., raising completion reduces ratings), change-points may appear in multiple KPI series with opposite signs, which itself can be informative for policy.

Nonlinearities, thresholds, and piece rates. Our maintained linearity assumption is not innocuous: many real contracts include thresholds, caps, streak bonuses, and tournaments. Two comments temper this limitation. First, an auditor can often locally linearize incentives around the realized outcome path: if payments are piecewise-linear in the audited KPI, then α_t can be interpreted as a *marginal* bonus over the region in which most observations fall. Second, the economic mechanism we exploit is not linearity per se but the presence of a low-dimensional statistic of past incentives that learners respond to (here $\bar{\alpha}_t$). Mean-based learning with recency weighting or reference dependence would replace $\bar{\alpha}_t$ with another moving average; the empirical signature would then be a different, estimable transformation (e.g., exponential decay rather than $1/t$). A useful extension is therefore to treat the "memory kernel" as a nuisance object and test whether $\hat{s}_t$ change-points align with *any* plausible discounted-average path generated by an abrupt downshift in incentives.

Stochastic stopping, churn, and unbalanced panels. Churn is central in many labor platforms and can be endogenous to incentives. We already allow for a per-period hazard h, but two distinct issues arise. The first is statistical: late-stage behavior under free-fall is observed on a selected sample of survivors, shrinking effective N_t and increasing the variance of $\hat{s}_t$. This does not invalidate our moment-inequality logic, but it does widen the slack $\delta_{N,T}$ and reduces power, especially for later breakpoint hits. Practically, auditors should (i) report N_t alongside $\hat{s}_t$, (ii) weight change-point detection by precision (e.g., approximate inverse-variance weights), and (iii) treat weak late-stage evidence as suggestive rather than dispositive.

The second issue is economic: if exit depends on α_t or on realized payoffs, the population composition may change even if each individual follows mean-based learning. Then a drop in $\hat{s}_t$ can reflect selection rather than effort reduction. One way forward is to model a joint learning–participation decision in which agents compare continuation value to an outside option. This produces an additional testable implication: under free-fall, exit hazards should rise around the same times that $\bar{\alpha}_t$ crosses indifference regions. Empirically, auditors can therefore run a "stacked" audit that checks whether breaks in $\hat{s}_t$ coincide with breaks in exit rates. A mismatch—large outcome change without a corresponding participation response, or vice versa—helps discriminate effort switching from compositional shifts.

Multi-agent interactions and endogenous success probabilities. In many applications, an agent's success probability is not time-invariant conditional on action because it depends on market thickness, congestion, matching algorithms, or peer behavior. Formally, q_a may become $q_a(x_t)$ for an aggregate state x_t that evolves endogenously with the population's actions. This creates two challenges. First, segmentation of $\hat{s}_t$ may detect change-points driven by demand shocks rather than effort. Second, even under free-fall incentives, the mapping from $\bar{\alpha}_t$ to outcomes may no longer be piecewise-constant if x_t drifts.

We see two tractable extensions. The first is a "slowly varying state" approximation: if x_t moves slowly relative to learning-induced switching, then breakpoint-hitting remains sharp while segment means drift within segments. Auditors can adapt by allowing within-segment trends and focusing the test on the *timing* of discrete drops rather than on level constancy. The second is to exploit quasi-experimental variation in x_t: when covariates capturing demand or congestion are observed, auditors can residualize outcomes and apply our protocol to the residual series. Importantly, our approach is conservative in the sense that it only delivers *necessary* conditions for free-fall compatibility; failure of the test is informative even in richer environments, while passing the test should be interpreted as "compatible with" rather than "proof of."

Regulatory and governance implications. From a policy perspective, our main contribution is to convert a qualitative concern—that platforms may temporarily use generous incentives to build habits and then reduce pay—into quantitative, falsifiable restrictions using only observables. This suggests three practical uses. First, regulators (or internal compliance teams) can deploy our audit protocol as a screening tool for markets in which incentive opacity is suspected, focusing investigative resources where multiple breakpoint hits line up with the free-fall invariants. Second, when concerns are substantiated, our surplus-transfer lower bounds provide a principled starting point for remedies (e.g., restitution or mandated disclosure), while explicitly accounting for partial identification and statistical uncertainty. Third, our results support targeted transparency mandates: requiring platforms to log and disclose marginal incentive schedules at a suitable granularity materially increases auditability because measurement error in α_t directly weakens the restrictions.

At the same time, we emphasize limitations that matter for enforcement. Our test is not a welfare test: free-fall compatibility does not imply consumer harm, anticompetitive conduct, or illegality; it indicates a particular dynamic incentive pattern consistent with surplus shifting from agents to the principal. Moreover, platforms may have benign reasons to change incentives (learning about demand, onboarding subsidies, seasonal policies). For this reason, an evidence-based regulatory approach should treat our compatibility assessment as one input into a broader factual record, ideally combined with contemporaneous product changes, demand measures, and stated policy rationales. The methodological goal is modest but useful: to make dynamic contracting claims auditable, reproducible, and quantitatively disciplined in settings where effort is inherently latent.