Dynamic Incentive Compatibility as Control: Online Monitoring and Stabilization via I-DIC

Liz Lemma Future Detective

January 16, 2026

Abstract

Most deployed auctions in 2026 are dynamic: bids today affect future eligibility, prices, pacing, quality scores, or platform recommendations. Classic DSIC/BIC are either inapplicable or too strong, and learning-based mechanisms (e.g., RegretNet-style) typically certify only static incentive properties. Building on the data-driven Stage-IC and Dynamic-IC metrics introduced in the source survey (I-SIC, I-DIC), we propose a control-theoretic framework for repeated auctions that treats dynamic incentive compatibility as a monitorable stability objective. We (i) formalize I-DIC as a local sensitivity of discounted utility to bid shading, (ii) give an online estimator using logged outcomes and small bid perturbations or counterfactual models, (iii) design a primal–dual controller that updates the mechanism to keep I-DIC below a target threshold, and (iv) prove a stability bound: if I-DIC $\leq \varepsilon$ each period, then the maximum discounted gain from any admissible shading policy is $O(\varepsilon/(1-\gamma))$. The result yields a practical alternative to full dynamic DSIC: platforms can enforce incentive stability with minimal assumptions, while trading it off against revenue in a transparent, tunable way. We validate the approach in ad-auction-like simulations and show reduced manipulation incentives under drift compared to static regret-minimization baselines. Numerical methods are needed only for the counterfactual utility estimation step when the mechanism is not fully observed/differentiable.

Table of Contents

3. 3. Model: repeated contextual auction with history dependence; bidder objectives; mechanism class; admissible deviations; definitions of stage and dynamic incentive stability.

4. 4. Metrics as sensitivity: formal definition of I-SIC and I-DIC; interpretation as local derivative of (discounted) utility; when metrics are well-defined (differentiability/Lipschitz assumptions).

5. 5. Estimation: online finite-difference estimator from logged outcomes; two settings—(a) differentiable/known mechanism enabling direct counterfactual evaluation; (b) black-box mechanism requiring model-based or importance-sampling off-policy evaluation (flag numerical requirements and variance control).

6. 6. Control algorithm: constrained revenue optimization with I-DIC caps; primal–dual updates; safety filters; practical implementation details (step sizes, smoothing, per-bidder vs aggregate constraints).

7. 7. Theory I (stability): metric-to-global-gain bound linking I-DIC to maximum discounted advantage of bid shading; extensions to multiple bidders and heterogeneous γ_i.

8. 8. Theory II (online performance): revenue regret bounds for the controller relative to best fixed feasible mechanism under convex surrogate assumptions; discussion of nonconvex reality and what is empirically validated.

9. 9. Experiments: ad-auction-like repeated setting (quality-score/pacing), cloud resource allocation, and/or mobility charging; distribution shift; comparison vs static regret minimization and no-control baselines; sensitivity to ε, γ, estimator noise.

10. 10. Discussion & limitations: admissible deviation class; dependence on Lipschitzness and counterfactual estimation; connections to dynamic DSIC; audit implications; open problems.

1 Introduction

Digital platforms increasingly allocate scarce opportunities through mechanisms that repeat, adapt, and learn. Sponsored-search systems run millions of auctions per day while continuously updating ranking and pricing rules; marketplaces adjust reserve prices, matching, and fee schedules in response to realized demand; and gig or creator platforms tune eligibility and exposure policies based on observed performance. In these environments, a bidder or seller is not merely choosing a one-shot bid. Rather, she is choosing an action today that both affects today's allocation and shapes tomorrow's mechanism through the platform's data-driven updates and the public history that competitors observe. This feedback loop is the defining feature of modern market design in practice, and it makes dynamic incentives first-order.

A central lesson from classical mechanism design is that dominant-strategy incentive compatibility (DSIC) provides a robust benchmark: if truthful reporting is a dominant strategy, we can reason about welfare and revenue without modeling strategic manipulation in detail. Yet DSIC is a static concept. Once the mechanism depends on history and is updated online, even mechanisms that are "truthful in each period" can create intertemporal incentives: shading a bid today may change future prices, eligibility, or inferred quality, thereby changing continuation payoffs. Conversely, mechanisms that are not exactly DSIC may still be acceptably robust in practice if the marginal gains from local deviations are small. This observation motivates our approach: rather than treating incentive compatibility as a binary property, we study and control incentive *sensitivity* in repeated, learning-driven auctions.

The limitations of static notions become especially stark when platforms pursue performance objectives via machine learning. Contemporary designs often parameterize allocation and payment rules by a high-dimensional vector and update it using stochastic gradient methods. Even if one could, in principle, enforce DSIC by construction, doing so may require restrictive functional forms or strong distributional assumptions that are at odds with the richness of real-world contexts. At the same time, standard online learning guarantees such as regret bounds typically treat the environment as exogenous. When agents are strategic and forward-looking, the data used for learning is itself an equilibrium object: the platform updates based on bids that respond to the update rule. Thus, the platform faces a joint problem of *learning* and *incentive management*, where small changes in mechanism parameters can create large changes in strategic behavior, and vice versa.

We propose a framework that makes this interaction measurable and, crucially, controllable. The key idea is to quantify the local profitability of bid shading around truthful bidding through a directional derivative of discounted utility with respect to a multiplicative perturbation. Intuitively, if a bidder slightly scales her bid by a factor $1 + \alpha$, the resulting change

in her expected current payoff and continuation value reveals how "steep" her objective is in the direction of manipulation. When this slope is near zero, truthful bidding is locally stable: the bidder may still be able to improve by large deviations, but small, feasible manipulations (which are often the relevant ones given operational constraints and uncertainty) yield little gain. This notion aligns with practice: platforms and auditors rarely need a theorem that *no* deviation is profitable; instead they need evidence that meaningful deviations are not profitable enough to justify engineering effort or compliance risk.

Our first contribution is to formalize two closely related sensitivity metrics: a *stage* incentive sensitivity (capturing the immediate effect of shading on the current round) and a *dynamic* incentive sensitivity (capturing the effect on current utility plus discounted continuation utility under a reference continuation behavior, such as truthful bidding thereafter). The dynamic metric directly targets the intertemporal channel that is central in repeated settings. Importantly, these objects can be estimated online using symmetric finite differences induced by small bid perturbations. This makes incentive monitoring feasible even when the mechanism is complex or learned, provided the platform can evaluate or approximate counterfactual outcomes under slightly perturbed bids. The resulting estimator has a transparent bias–variance tradeoff governed by the perturbation size, and its interpretation is straightforward: it is a normalized "marginal gain from shading" measured in units of truthful expected surplus.

Our second contribution is conceptual: we show how bounding dynamic incentive sensitivity translates into a bound on the total discounted gains from (restricted) strategic manipulation. The economic logic is simple. If the bidder's discounted value function is locally flat in a neighborhood around truthful bidding—in the sense that its derivative with respect to shading is uniformly small—then even an optimally chosen, history-dependent shading policy cannot accumulate large benefits. Discounting plays a central role: a per-period incentive slope can compound over time, but the geometric discount factor limits this amplification, yielding bounds that scale like $1/(1-\gamma)$. This makes precise a tradeoff that platform designers already face implicitly. Tighter incentive stability (smaller allowable sensitivity) reduces opportunities for manipulation but restricts the designer's ability to extract revenue via aggressive pricing or history dependence. Our results make this tradeoff explicit and provide constants that tie it to primitives such as Lipschitz continuity of per-round utility in bids and the allowed magnitude of shading.

Our third contribution is methodological and speaks directly to implementation. We propose an online primal–dual control rule that treats incentive sensitivity as a constraint, with a Lagrange multiplier that adapts in real time. The platform updates its mechanism parameters to increase an objective such as revenue, while simultaneously penalizing violations of the

incentive sensitivity budget. This is a natural fit for platforms already operating gradient-based pipelines: the same machinery used to optimize performance can incorporate an incentive "risk" signal. Under standard convexity and bounded-gradient conditions (on suitable surrogates), the resulting algorithm achieves sublinear regret relative to the best fixed design satisfying the sensitivity constraint, while ensuring that average constraint violation vanishes. In economic terms, we obtain a disciplined way to run a revenue-optimizing mechanism *subject to* an auditable notion of incentive robustness.

The framework also clarifies what can and cannot be guaranteed. Our sensitivity metrics are local by construction; they certify stability against small bid shading around truthful behavior, not global optimality across arbitrary deviations. This is a feature rather than a bug when the designer's goal is operational robustness, but it is a limitation for settings where agents can costlessly implement complex deviations. Likewise, the quality of incentive control depends on the quality of counterfactual evaluation: if the platform's model of perturbed outcomes is biased, sensitivity may be under- or over-estimated, leading respectively to manipulability or excessive conservatism. These considerations suggest a natural role for monitoring and auditing: an external party can verify the measurement pipeline and impose a cap on allowable dynamic incentive sensitivity, analogous to how risk limits are imposed in safety-critical systems.

Finally, we emphasize the practical interpretation of our approach. Rather than asking whether a learned auction is exactly truthful—a demanding and often brittle requirement—we ask whether it is *approximately stable* in the sense that marginal incentives to shade are small throughout its operation. This shift from exact DSIC to controlled incentive sensitivity is well-suited to dynamic platforms, where mechanisms evolve, contexts are high-dimensional, and the relevant threats are incremental manipulations that exploit predictable gradients in pricing or ranking rules.

The remainder of the paper develops these ideas as follows. We first situate our work within the literatures on learning-based mechanism design, dynamic auctions, and online control of economic systems. We then define the stage and dynamic incentive sensitivity metrics and show how to estimate them online with finite differences, including a discussion of tuning and robustness. Next, we establish bounds connecting sensitivity control to bounded strategic gains under restricted deviation classes, highlighting the role of discounting and regularity. We then present the primal–dual controller and analyze its regret and constraint-violation guarantees. We conclude with implications for platform governance and open questions on extending local certification to richer deviation models and weaker counterfactual assumptions.

2 Related Work

Our work sits at the intersection of learning-based mechanism design, dynamic and repeated auctions, and online control in strategic environments. A useful organizing theme in this literature is the tension between *expressiveness* (mechanisms parameterized by rich function classes and tuned from data) and *incentive guarantees* (whether truthful or near-truthful behavior can be expected when agents are forward-looking). We view our contribution as providing a measurement-and-control layer that can be attached to a broad class of learned, history-dependent mechanisms: rather than insisting on exact incentive compatibility, we monitor and bound the *marginal* profitability of manipulation in the directions that are operationally salient.

Learning auctions with regret-based objectives. A prominent approach to automated mechanism design trains a parameterized allocation and payment rule by minimizing an empirical notion of *regret*, often alongside a revenue or welfare objective. RegretNet and related architectures **?** implement this idea by sampling value profiles, optimizing payments and allocations by gradient descent, and penalizing deviations from truthful reporting measured by the best-response utility gain within the sample. This line of work has two features that are particularly relevant for us. First, it highlights the practical appeal of incentive constraints expressed as *loss terms* that can be optimized with standard ML tooling. Second, it makes clear that, even in static settings, incentive compatibility is typically enforced approximately and empirically, with guarantees that depend on the richness of the deviation class searched and the quality of best-response computation. Our focus differs in that we target repeated environments in which the mechanism depends on public history and may update online. In such settings, the incentive problem is inherently intertemporal: the relevant deviations alter not only current outcomes but also future states through learning and feedback. Regret-style penalties can be extended to dynamic contexts, but doing so requires specifying a dynamic deviation model and solving a dynamic best-response problem, which is computationally and statistically demanding.

Architectures enforcing structure: RochetNet and MyersonNet. A complementary line of work builds incentive properties into the parameterization itself. RochetNet **?** and related methods exploit Rochet's characterization of implementable allocation rules via convex potentials in quasi-linear environments, ensuring incentive compatibility by construction (often up to approximation error). MyersonNet **?** and subsequent work similarly embed Myerson's virtual-value logic, learning monotone transformations or reserve policies aligned with revenue-optimality under distributional assumptions. These approaches underscore an important design philosophy: if we

can encode the right structural constraints (monotonicity, convexity, envelope conditions), we can obtain strong, global incentive guarantees. At the same time, the cost is reduced flexibility and, in many applied domains, a mismatch between the clean static models that admit exact characterizations and the messy, contextual, and stateful mechanisms used in practice. Our perspective is that, once mechanisms become history-dependent and are updated online, insisting on exact implementability may either be infeasible or may rule out useful classes of mechanisms (e.g., those incorporating nontrivial exploration or stateful eligibility rules). In such cases, a disciplined *approximate* notion of incentive robustness that remains measurable online becomes valuable.

Dynamic auctions and dynamic mechanism design. The classical theory of dynamic mechanism design studies environments in which private information evolves and the designer may condition on past reports and allocations. Seminal contributions characterize efficient and revenue-optimal dynamic mechanisms under various informational and commitment assumptions; see, among many others, **???**. This literature provides the conceptual foundation for understanding continuation values and intertemporal incentive constraints. However, optimal dynamic mechanisms are typically complex even under stylized assumptions, and the resulting prescriptions can be fragile when the environment is misspecified or when the platform must learn from interaction data. In repeated auctions run by platforms, history dependence is often introduced not to implement a theoretically optimal dynamic mechanism, but because the platform updates ranking, pricing, or eligibility rules based on observed outcomes, which in turn are influenced by strategic behavior. Our analysis is motivated by this operational reality: the intertemporal channel arises endogenously from learning and feedback rather than from a fully solved dynamic mechanism design problem.

Approximate incentive compatibility and local notions. A broad set of papers study relaxations of incentive compatibility, including additive and multiplicative approximate IC, ex post and interim notions, and empirical IC guarantees. Regret-based metrics used in learned auctions are one prominent example, but the idea of quantifying "how far" a mechanism is from IC has older roots in both mechanism design and econometrics. Our stage and dynamic sensitivity metrics, I-SIC and I-DIC, are most closely related to *local* or *first-order* relaxations: they measure the directional derivative of (discounted) utility with respect to a small, multiplicative shading of the truthful bid. Conceptually, this resembles sensitivity analysis in optimization and influence-function logic in statistics: we ask how the objective responds to an infinitesimal perturbation, normalized by a scale factor that makes magnitudes comparable across contexts. The key distinction is that we in-

corporate continuation value into the derivative, thereby directly targeting the dynamic feedback loop that is absent in static metrics. The limitation of any local metric is also clear: local flatness does not rule out profitable large deviations. We view this as an acceptable tradeoff in settings where deviations are practically constrained (by engineering costs, uncertainty, platform monitoring, or bid granularity) and where policy goals emphasize *robustness* rather than exact dominant strategies.

Online learning and control with constraints in markets. Our primal–dual controller builds on the extensive literature on online convex optimization and constrained online learning **??**, where one optimizes a sequence of objectives while keeping long-run constraint violations small via Lagrange multipliers. Similar ideas appear in "safe" learning and constrained Markov decision processes, where one trades off performance and risk through dual variables. In market settings, online learning techniques have been applied to dynamic pricing, bandit auctions, and adaptive reserve selection, often under assumptions that the environment is exogenous or that buyers are myopic. When agents are strategic and forward-looking, the platform's data is equilibrium-dependent, complicating standard regret interpretations. We do not attempt to solve this general strategic learning problem. Instead, we propose to treat incentive sensitivity as a monitored constraint, thereby creating a feedback mechanism that penalizes designs that generate steep manipulation gradients. This aligns with how platforms often operate: they may accept that strategic effects exist, but they seek operational guardrails that limit their magnitude.

Governance, auditing, and practical implementation. Finally, our framing connects to emerging discussions about auditing algorithmic marketplaces. In practice, regulators or internal risk teams rarely certify that a mechanism is exactly DSIC, especially when it is updated continuously. A more realistic governance question is whether the platform can demonstrate that the mechanism does not create *strong* incentives for predictable manipulation, particularly through small and systematic bid adjustments that sophisticated participants can automate. By casting this as a measurable sensitivity budget ε and embedding it into an online control loop, we provide a language for such audits that is analogous to risk limits in safety-critical systems. This said, the approach is only as credible as the counterfactual evaluation used to estimate I-DIC; model error or limited experimentation can bias measurement. Recognizing this limitation is essential: the contribution is not a substitute for structural mechanism design, but a pragmatic complement that helps manage incentives when rich, learned, and history-dependent mechanisms are unavoidable.

3 Model: repeated contextual auctions with history dependence

Environment and timing. We study a repeated auction run over a finite horizon T with n strategic bidders. In each round $t \in \{1, \dots, T\}$, a public context x_t is observed. The context can encode item features, quality scores, eligibility constraints, or market conditions that affect either the feasible allocations or the platform's ranking and pricing logic. A public history h_{t-1} summarizes all past publicly observed information up to $t-1$ (e.g., contexts, bid profiles, allocations, payments, and any platform announcements). Bidder i privately observes a per-round value $v_{i,t} \in [0,1]$ and submits a bid $b_{i,t}$. The platform then applies a (possibly history-dependent) mechanism to produce an allocation vector $a_t = (a_{1,t}, \dots, a_{n,t})$ and payments $p_t = (p_{1,t}, \dots, p_{n,t})$. The realized round-$t$ utility is quasilinear,

$$u_{i,t} \;=\; v_{i,t}\, a_{i,t} \;-\; p_{i,t}.$$

The history updates to h_t, and the process repeats.

Two features are central for our purposes. First, allocations and payments may depend on (x_t, h_{t-1}) as well as the full bid profile b_t; this captures a broad class of contextual and stateful marketplace rules. Second, the platform may update its mechanism over time using interaction data. We model this by a parameter θ_t and a mechanism family $\mathcal{M}_\theta$ such that, in round t,

$$(a_t, p_t) \;=\; \mathcal{M}_{\theta_t}(x_t, h_{t-1}, b_t),$$

with θ_{t+1} allowed to depend on the observed outcomes. This encompasses common operational patterns such as learning-to-rank updates, adaptive reserve policies, eligibility tuning, and budget pacing logic. Precisely because θ may evolve endogenously, a bidder's current bid can influence future allocation and pricing conditions through the public state.

Bidder objectives and continuation values. Bidders are forward-looking with discount factor $\gamma \in [0,1)$. For a (possibly history-dependent) bid policy π_i mapping the bidder's information into bids, bidder i's discounted utility is

$$U_i(\pi_i; \theta) \;=\; \mathbb{E}\left[\sum_{t=1}^{T} \gamma^{t-1} u_{i,t}\right],$$

where the expectation is over values, contexts, and any mechanism or platform randomness, induced by the joint policy profile and the mechanism's evolution. It is often convenient to separate the current-round effect of a bid from its impact on the future through the history. We therefore write a continuation value from time $t+1$ onward, conditional on the post-round

history h_t,

$$\bar{U}_{i,t}(h_t;\theta,\pi_i) \;=\; \mathbb{E}\left[\sum_{\tau=t+1}^{T}\gamma^{\tau-t}u_{i,\tau}\,\middle|\,h_t\right].$$

This quantity is the formal object through which the intertemporal incentive problem enters: even if a bid deviation is unprofitable myopically, it can be profitable dynamically if it steers the platform's state (or rivals' behavior) in a favorable direction.

Mechanism class and observability. We intentionally allow $\mathcal{M}_\theta$ to be rich and history dependent. From a theoretical perspective, this generality is what makes exact incentive compatibility hard to ensure: the mapping from a single bidder's bid to their overall discounted utility can become complex once learning, feedback, and state updates are present. From a practical perspective, however, it reflects how real platforms operate. A platform typically does not commit to a fixed auction rule for long horizons; instead it iterates, monitors, and updates. Our goal is therefore not to impose a narrow structural form, but to develop incentive notions that can be monitored in such adaptive regimes.

We also emphasize an informational asymmetry that matters for implementation: the platform generally observes bids, allocations, and payments, but may not observe values. For governance purposes, the relevant question is whether the platform can nonetheless detect and limit the profitability of systematic bid manipulations using the signals it does observe (possibly augmented by experiments or counterfactual estimation).

Dynamic best responses and the truthful benchmark. Fix a platform update rule (or a realized path of parameters $\theta_1,\ldots,\theta_T$). In round t, bidder i's dynamic best response trades off current utility and continuation value. Formally, letting $\hat{u}_{i,t}(b_{i,t};h_{t-1},v_{i,t})$ denote bidder i's interim per-round utility when they submit $b_{i,t}$ (holding $(h_{t-1},x_t,v_{i,t})$ fixed and integrating over other bidders' bids and any mechanism randomness), a dynamic best response solves

$$b_{i,t}\in\arg\max_{b\in\mathcal{B}}\;\mathbb{E}\left[\hat{u}_{i,t}(b;h_{t-1},v_{i,t})\;+\;\bar{U}_{i,t}(h_t;\theta,\pi_i)\,\middle|\,h_{t-1},v_{i,t}\right].$$

The canonical benchmark is truthful bidding, $b_{i,t}=v_{i,t}$, interpreted as the reporting policy the platform would like to elicit. An "ideal" dynamic incentive compatibility requirement would assert that truthful bidding is optimal at every history and value realization. In repeated, history-dependent environments, this requirement is typically too strong: it can fail even for mechanisms that are myopically truthful because the continuation value $\bar{U}_{i,t}$ introduces new strategic channels.

Admissible deviations and why we focus on shading. Rather than attempting to characterize all possible dynamic deviations, we adopt a deviation class that is both economically meaningful and operationally salient: multiplicative bid shading around truth. Concretely, we consider policies of the form

$$b_{i,t} \;=\; s_{i,t}(h_{t-1})\, v_{i,t},$$

with shading factors constrained to lie in a neighborhood of one, $s_{i,t}(h_{t-1}) \in [1-\bar{\alpha},\, 1+\bar{\alpha}]$ for some small $\bar{\alpha} > 0$. This class captures the kind of manipulation that sophisticated participants can automate at scale (e.g., uniform bid multipliers, pacing-like behavior, or systematic underbidding in response to perceived overpricing) without requiring them to solve a full dynamic program.

We view the restriction to local shading deviations as a deliberate tradeoff. It yields a tractable and monitorable notion of incentive robustness, but it does not preclude profitable large deviations or more complex misreports. This limitation is important in principle; nonetheless, in many marketplace settings the most common strategic behaviors are incremental and systematic, and platforms often seek guardrails that prevent "small hacks" from compounding into substantial advantage.

Stage versus dynamic incentive stability. The key conceptual distinction in a stateful mechanism is between incentives that operate within a round and incentives that operate through the state. To formalize this, we separate two objects:

(i) *Stage incentives* ask: holding the public state fixed and abstracting from future consequences, how profitable is a small deviation from truthful bidding in the current round? This corresponds to the familiar static IC logic applied conditional on (x_t, h_{t-1}).

(ii) *Dynamic incentives* ask: accounting for how current actions affect the next public history and thereby future allocations, payments, and platform updates, how profitable is a small deviation? This is the relevant notion when θ is updated online or when future eligibility/ranking depends on observed bids and outcomes.

Our aim is not to enforce exact optimality of truth-telling, but to bound the *marginal* gain from such deviations. Intuitively, we would like the bidder's discounted value as a function of the shading factor s to be locally flat at $s = 1$: if the slope is small, then nearby systematic shading provides little advantage, and the platform is less exposed to predictable forms of manipulation.

Local stability budgets as design constraints. We operationalize these ideas by introducing per-round *incentive stability* metrics: one for the stage effect and one for the dynamic (stage plus continuation) effect. Each metric

is designed to be (i) local, in the sense of focusing on a small perturbation around truthful bidding; (ii) normalized, so that magnitudes are comparable across contexts; and (iii) amenable to online estimation using bid perturbations or counterfactual evaluation. A platform (or an external auditor) can then impose a stability budget ε, requiring that the local profitability of shading remain below ε each round (or on average conditional on history). This creates a concrete governance handle: rather than certifying global IC in an adaptive system, the platform can demonstrate that it keeps the manipulation gradient small in the directions most likely to be exploited.

In the next section we formalize these metrics as directional derivatives of (discounted) utility with respect to the shading factor and state the regularity conditions under which they are well-defined and computable.

4 Metrics as sensitivity: stage and dynamic slopes around truth

We now formalize the idea that an adaptive, history-dependent mechanism can be *locally* robust to manipulation even when it is not globally (dynamic) incentive compatible. The object we can hope to monitor in real time is not a global deviation gain—which depends on a high-dimensional policy class—but rather the *marginal* gain from systematic shading in a neighborhood of truthful bidding. Concretely, we perturb bidder i's truthful bid $v_{i,t}$ by a multiplicative factor s close to one and ask how the bidder's (discounted) value changes at $s = 1$.

Directional perturbations and the stage utility map. Fix a round t, a public history h_{t-1}, and bidder i's value realization $v \in [0, 1]$. We write $\hat{u}_{i,t}(b; h_{t-1}, v)$ for bidder i's interim stage utility when they submit bid b, holding (h_{t-1}, x_t, v) fixed and integrating over rivals' bids and any mechanism randomness induced by $\mathcal{M}_{\theta_t}$. We then consider the one-dimensional bid path

$$b \;=\; sv, \qquad s \in \mathbb{R}_+,$$

and focus on local deviations $s = 1 + \alpha$ for small α. The multiplicative parameterization is convenient for two reasons. First, it aligns with common operational manipulations (uniform bid multipliers, conservative underbidding, pacing-style attenuation). Second, it is scale-consistent in the sense that a fixed α represents the same proportional change for low and high values.

Stage incentive stability: I-SIC. The stage metric measures the local slope of *current-round* utility with respect to shading, abstracting from any

effect of the bid on the future state. Formally, whenever $\hat{u}_{i,t}(sv; h_{t-1}, v)$ is differentiable in s at $s = 1$, we define the (signed) stage sensitivity

$$\text{I-SIC}_{i,t} \;:=\; \frac{\frac{\partial}{\partial \alpha}\,\mathbb{E}\big[\hat{u}_{i,t}\big((1+\alpha)v_{i,t}; h_{t-1}, v_{i,t}\big)\,\big|\,h_{t-1}\big]\big|_{\alpha=0}}{\mathbb{E}\big[v_{i,t}\,a_{i,t}(v_{i,t})\,\big|\,h_{t-1}\big]}.$$

The numerator is the directional derivative at truth in the "shading direction" $b = (1+\alpha)v$. The denominator normalizes by a truthful benchmark scale, $\mathbb{E}[va(v) \mid h_{t-1}]$, which can be interpreted as expected truthful gross value (welfare for bidder i) in round t. This normalization makes the metric comparable across contexts and histories with very different levels of trade or assignment probability. (When $\mathbb{E}[va(v) \mid h_{t-1}] = 0$, the round is effectively irrelevant for bidder i, and the metric is either undefined or taken to be 0 by convention; in what follows we assume it is bounded away from zero on the histories of interest.)

The sign of $\text{I-SIC}_{i,t}$ is informative: a positive value indicates that locally increasing the bid (relative to value) increases interim stage utility, whereas a negative value indicates local profitability of shading down. For governance, we typically care about the *magnitude* of local manipulability, and thus impose a budget on $|\text{I-SIC}_{i,t}|$ (or, in one-sided formulations, on the positive part if only overbidding is relevant). A small bound $|\text{I-SIC}_{i,t}| \le \varepsilon$ implies that for small α,

$$\mathbb{E}\big[\hat{u}_{i,t}\big((1+\alpha)v_{i,t}\big) - \hat{u}_{i,t}(v_{i,t})\,\big|\,h_{t-1}\big] \;\approx\; \alpha \cdot \text{I-SIC}_{i,t} \cdot \mathbb{E}\big[v_{i,t}a_{i,t}(v_{i,t})\,\big|\,h_{t-1}\big]\,,$$

so the dollar gain from small systematic shading is proportional to ε times the truthful scale.

Dynamic incentive stability: I-DIC. In a stateful mechanism, a bid can also change the next history h_t and thereby the continuation value. We therefore extend the same sensitivity logic to *discounted* utility. Fixing a reference continuation behavior (e.g., truthful bidding from $t+1$ onward) and the platform's update rule for θ, let $\bar{U}_{i,t}(h_t; \theta, \pi_i)$ denote bidder i's continuation value from $t+1$ onward as defined earlier. We view $\bar{U}_{i,t}$ as a function of the post-round history h_t, and hence (implicitly) as a function of the current bid through the induced distribution over h_t. When this composite map is differentiable along the shading path at $s = 1$, we define

$$\text{I-DIC}_{i,t} \;:=\; \frac{\frac{\partial}{\partial \alpha}\,\mathbb{E}\big[\hat{u}_{i,t}\big((1+\alpha)v_{i,t}\big) \;+\; \bar{U}_{i,t}\big((1+\alpha)v_{i,t}\big)\,\big|\,h_{t-1}\big]\big|_{\alpha=0}}{\mathbb{E}\big[v_{i,t}\,a_{i,t}(v_{i,t})\,\big|\,h_{t-1}\big]}.$$

This metric captures the full local incentive effect of shading: the immediate change in payments/allocations *plus* the induced change in the distribution of future states (and thus future utilities). It is therefore the relevant constraint

when θ is updated online, when eligibility depends on past bids/outcomes, or when any part of the mechanism is explicitly history dependent.

As with the stage metric, the dynamic metric can be read as a local first-order condition: if the bidder's discounted objective is locally maximized at truthful bidding along the shading direction, then the derivative is zero (or satisfies a suitable inequality at the boundary), implying $\text{I-DIC}_{i,t} \approx 0$. More generally, requiring $|\text{I-DIC}_{i,t}| \leq \varepsilon$ enforces that the discounted value function is locally flat at $s = 1$, so that no small multiplicative shading rule can reliably extract a large advantage through either the current mechanism mapping or the induced state dynamics.

When are the metrics well-defined? The definitions above are local derivatives, so they require mild regularity near truthful bidding. Our standing assumption (H1) ensures that for each (i, t), the interim stage utility $\hat{u}_{i,t}(b; h_{t-1}, v)$ is differentiable in b at $b = v$ and L-Lipschitz in b uniformly over (h_{t-1}, v). Lipschitzness serves two roles. First, it rules out pathological sensitivity: a small bid perturbation cannot create an arbitrarily large utility jump. Second, it provides the control needed for finite-difference approximations: for α small, the deviation payoff is approximately linear in α, with remainder terms controlled by smoothness (or, more weakly, by generalized derivative bounds). The boundedness assumption (H2) similarly ensures that continuation values are well-defined and that interchanging differentiation and expectation is justified under standard dominated convergence arguments.

It is worth emphasizing a practical point: many allocation rules are not pointwise differentiable in bids (e.g., deterministic winner-take-all rules). What matters here is differentiability of the *interim* objective $\hat{u}_{i,t}$ along the shading path after integrating over other bidders and mechanism randomness. Random tie-breaking, reserve noise, smoothing in ranking, or context variability can all render the interim map differentiable even when the realized allocation is discontinuous. When differentiability fails, Lipschitzness still implies almost-everywhere differentiability (by Rademacher's theorem) and supports interpreting the metric via directional derivatives or subgradients; empirically, our estimators in the next section target the same local slope through symmetric perturbations.

A note on normalization and interpretation. Finally, the normalization by $\mathbb{E}[va(v) \mid h_{t-1}]$ is not merely cosmetic. Without it, the same absolute derivative could correspond to negligible strategic significance in low-volume contexts and major manipulability in high-volume contexts. The normalized metric expresses the marginal gain from shading as a fraction of a truthful-scale benchmark, aligning with the idea of a platform-wide stability budget ε that is meaningful across heterogeneous rounds and bidder populations.

Having defined I-SIC and I-DIC as local slopes, we next turn to how a platform (or auditor) can estimate them online from logged interaction data, both when the mechanism admits direct counterfactual evaluation and when it must be treated as a black box.

5 Online estimation of local incentive slopes from logged interaction

The definitions of I-SIC$_{i,t}$ and I-DIC$_{i,t}$ are directional derivatives, so the natural empirical counterpart is a finite-difference slope computed around the truthful point $b = v$. The main implementation question is counterfactual access: can we evaluate what would have happened in round t had bidder i submitted $(1 \pm \alpha)v_{i,t}$ while all else in that round was held fixed? We describe two settings that cover most platform deployments.

A symmetric finite-difference template. Fix (i, t) and a history h_{t-1}. Consider two perturbed bids,

$$b^+_{i,t} := (1 + \alpha)v_{i,t}, \qquad b^-_{i,t} := (1 - \alpha)v_{i,t},$$

with $\alpha > 0$ small. Let $W_{i,t}(b)$ denote the bidder's discounted objective "starting in round t" evaluated at bid b in round t and a specified continuation rule thereafter:

$$W_{i,t}(b) \;:=\; \hat{u}_{i,t}(b; h_{t-1}, v_{i,t}) \;+\; \bar{U}_{i,t}(b; h_{t-1}, v_{i,t}).$$

A direct finite-difference estimate of the (unnormalized) derivative is

$$\widehat{D}_{i,t} \;:=\; \frac{\widehat{W}_{i,t}(b^+_{i,t}) - \widehat{W}_{i,t}(b^-_{i,t})}{2\alpha},$$

and the corresponding normalized metric estimate is

$$\widehat{\text{I-DIC}}_{i,t} \;:=\; \frac{\widehat{D}_{i,t}}{\widehat{Z}_{i,t}}, \qquad \widehat{Z}_{i,t} \approx \mathbb{E}\big[v_{i,t}a_{i,t}(v_{i,t}) \,\big|\, h_{t-1}\big]\,.$$

Using the symmetric difference is important in practice: under smoothness, it has bias $O(\alpha^2)$, while the one-sided difference has bias $O(\alpha)$. The same template applies to I-SIC by dropping $\bar{U}_{i,t}$.

Two operational details matter. First, the denominator $\widehat{Z}_{i,t}$ is itself an object of conditional expectation; a single realization $v_{i,t}a_{i,t}$ can be too noisy when trade is sparse. In applications we typically use a rolling regression or moving average (conditioning on coarse bins of x_t and h_{t-1}) so that the normalization does not spuriously explode. Second, because $\widehat{D}_{i,t}$ is computed per-round, platforms usually smooth $\widehat{\text{I-DIC}}_{i,t}$ across time (e.g., an exponential moving average) before feeding it into a controller, trading responsiveness for variance reduction.

(a) Known and (interim) differentiable mechanisms: direct counterfactual evaluation. In the most transparent setting, the platform knows $\mathcal{M}_{\theta_t}$ and can re-run the allocation and payment mapping on the realized bid profile with bidder i's bid replaced. Concretely, in round t we observe (x_t, h_{t-1}, b_t) and the mechanism's internal randomness (or its random seed). We can then compute counterfactual outcomes

$$(a_t^+, p_t^+) = \mathcal{M}_{\theta_t}\big(x_t, h_{t-1}, (b_{i,t}^+, b_{-i,t})\big), \qquad (a_t^-, p_t^-) = \mathcal{M}_{\theta_t}\big(x_t, h_{t-1}, (b_{i,t}^-, b_{-i,t})\big),$$

and form stage-utility realizations

$$\widehat{u}_{i,t}(b_{i,t}^{\pm}) \;:=\; v_{i,t}\, a_{i,t}^{\pm} - p_{i,t}^{\pm},$$

where $v_{i,t}$ is treated as the value at which the directional derivative is defined (in a lab setting it is known; in the field, one can replace it with a calibrated proxy or focus on sensitivity with respect to the reported bid itself). The remaining term is the continuation value. A practically useful representation is a value-function approximation $\widehat{V}_{i,t}(\cdot)$ mapping post-round public history to expected discounted future utility under the reference continuation rule:

$$\widehat{\widetilde{U}}_{i,t}(b) \;\approx\; \gamma\, \widehat{V}_{i,t}(h_t(b))\,,$$

where $h_t(b)$ is the counterfactual next history induced by submitting b in round t (including the counterfactual allocation/payment and any public state variables updated from them). This turns the dynamic slope into a one-step lookahead object, analogous to temporal-difference methods in reinforcement learning, and it avoids high-variance full-horizon rollouts. When the horizon is short (or when an accurate simulator is available), Monte Carlo rollouts under the reference continuation rule are also feasible:

$$\widehat{\widetilde{U}}_{i,t}(b) \;=\; \sum_{\tau=t+1}^{t+H} \gamma^{\tau-t}\, \widehat{u}_{i,\tau}(b),$$

with truncation H controlling variance and bias.

Direct counterfactual evaluation is numerically stable because it does not require importance weights; the main tuning knob is α. Too large an α invalidates the local approximation; too small an α makes the difference $\widehat{W}(b^+) - \widehat{W}(b^-)$ dominated by noise from discrete allocation changes and value-function error. In deployments we therefore choose α by monitoring the empirical signal-to-noise ratio of the numerator and by imposing an α-floor to avoid numerical cancellation.

(b) Black-box mechanisms: model-based and off-policy estimators. When the platform cannot reliably re-run $\mathcal{M}_{\theta_t}$ on arbitrary counterfactual bids (e.g., due to proprietary components, non-deterministic downstream

systems, or missing internal randomness), $\widehat{W}_{i,t}(b^{\pm}_{i,t})$ must be estimated from logged data. This is an off-policy evaluation (OPE) problem: we observed outcomes under the behavior distribution of bids, but we need the value under a nearby "target action".

A model-based approach learns a differentiable surrogate $\widehat{\mathcal{M}}_\phi$ from logs, mapping (x_t, h_{t-1}, b_t) to predicted allocations and payments. One then evaluates $\widehat{\mathcal{M}}_\phi$ at $(b^{\pm}_{i,t}, b_{-i,t})$ to obtain $(\widehat{a}^{\pm}_t, \widehat{p}^{\pm}_t)$, and proceeds exactly as in case (a), replacing true counterfactual outcomes with predicted ones. The advantage is low variance; the limitation is bias from model misspecification, which directly feeds into $\widehat{\text{I-DIC}}$ and can cause systematic under-enforcement of the constraint. For this reason, we prefer flexible models with calibrated uncertainty, and we recommend periodic backtests against any available partial ground truth (e.g., replay in controlled sandboxes).

Alternatively, one can use importance sampling when the platform knows (or can estimate) the conditional bid density $\mu_{i,t}(b \mid h_{t-1}, v_{i,t})$ that generated observed bids. For a stochastic target policy $\pi^{\pm}_{i,t}$ concentrated near $b^{\pm}_{i,t}$, we can form weights

$$w^{\pm}_{i,t} \;:=\; \frac{\pi^{\pm}_{i,t}(b_{i,t} \mid h_{t-1}, v_{i,t})}{\mu_{i,t}(b_{i,t} \mid h_{t-1}, v_{i,t})},$$

and estimate $\mathbb{E}[W_{i,t}(b^{\pm}_{i,t})]$ by a weighted average of realized $W_{i,t}(b_{i,t})$. Two numerical requirements are non-negotiable: *overlap* (the behavior policy must put sufficient mass where the target policy concentrates) and *weight control* (the ratio must not have heavy tails). In practice this pushes us toward (i) deliberately randomized exploration that occasionally perturbs bids or bid multipliers so that nearby actions have support, and (ii) variance reduction via clipped weights, self-normalized importance sampling, and/or doubly robust estimators that combine a reward model with importance weights:

$$\widehat{W}^{\mathrm{DR}} \;=\; \widehat{W}^{\mathrm{model}}(b^{\pm}) \;+\; w^{\pm}\Big(W(b) - \widehat{W}^{\mathrm{model}}(b)\Big).$$

The doubly robust form is particularly attractive here because the incentive-slope estimator is a *difference of two nearby counterfactual values*; correlated errors can cancel if the model is smooth, while the importance term guards against bias when overlap is adequate.

A final practical limitation is that deterministic targets $b = b^{\pm}_{i,t}$ are incompatible with importance sampling unless the behavior has point mass at exactly those bids. We therefore interpret the derivative operationally as the slope of a *smoothed* objective, using a narrow stochastic target around $(1 \pm \alpha)v$ (e.g., log-normal noise on the multiplier). This makes the OPE problem well-posed and aligns with the fact that, empirically, we care about systematic shading rules with small but nonzero dispersion.

Taken together, these estimators provide a feasible monitoring layer: when counterfactual replay is available, $\widehat{\text{I-DIC}}$ can be computed with relatively modest statistical complexity; when it is not, the platform must either

invest in credible surrogate models or introduce controlled randomization to ensure overlap, accepting the accompanying bias–variance and governance tradeoffs.

6 Online control: constrained revenue optimization with I-DIC caps

Once $\widehat{\text{I-DIC}}_{i,t}$ is available as a monitoring signal, the remaining design problem is operational: how do we *adapt* the mechanism parameters θ to improve revenue while ensuring that local dynamic incentives remain within a prescribed stability budget ε? We treat this as a constrained online optimization problem in which each round produces (i) a noisy revenue gradient signal and (ii) a noisy constraint signal based on incentive slopes. The controller then updates θ in small increments, with explicit safeguards to prevent transient constraint blow-ups from propagating into large regime shifts.

A Lagrangian viewpoint and primal–dual updates. Let $R_t(\theta)$ denote a per-round revenue objective (possibly discounted by the platform) evaluated under $\mathcal{M}_\theta$ at the realized context and bid profile, and let $G_{i,t}(\theta)$ denote the round-t incentive-sensitivity quantity we aim to cap, e.g.

$$G_{i,t}(\theta) \;:=\; \text{I-DIC}_{i,t}(\theta), \qquad \text{with constraint} \qquad G_{i,t}(\theta) \le \varepsilon.$$

In many deployments θ affects reserves, rank-score coefficients, pacing parameters, or other pricing rules, and $R_t(\theta)$ is not literally concave. Nonetheless, a robust control template is to run a primal–dual update on a surrogate Lagrangian with stochastic gradients:

$$\mathcal{L}_t(\theta, \lambda_t) \;=\; \widehat{R}_t(\theta) \;-\; \sum_{i=1}^{n} \lambda_{i,t}\big(\widehat{G}_{i,t}(\theta) - \varepsilon\big), \qquad \lambda_{i,t} \ge 0,$$

with a primal step that moves θ in the direction of higher revenue penalized by constraint pressure, and a dual step that increases penalties when the estimated incentive slope exceeds ε. Concretely, for step sizes $\eta_\theta, \eta_\lambda > 0$ and a feasible parameter domain Θ (e.g. box constraints), we implement

$$\theta_{t+1} = \Pi_\Theta\left(\theta_t + \eta_\theta\Big(\nabla_\theta \widehat{R}_t(\theta_t) - \sum_{i=1}^{n} \lambda_{i,t} \nabla_\theta \widehat{G}_{i,t}(\theta_t)\Big)\right), \qquad \lambda_{i,t+1} = \Big[\lambda_{i,t} + \eta_\lambda\big(\widehat{G}_{i,t}(\theta_t) - \varepsilon\big)\Big]_+,$$

where Π_Θ is projection and $[\cdot]_+$ is truncation at zero. Even when gradients are only approximate (e.g. obtained from a differentiable surrogate, or via bandit feedback), the structure is useful: the dual variables $\lambda_{i,t}$ become interpretable "prices of manipulability," rising precisely when the monitored incentive slope drifts upward, and relaxing when the mechanism is locally flat around truthful bidding.

Per-bidder versus aggregate constraints. A practical decision is whether to maintain n separate constraints $G_{i,t} \le \varepsilon$ or to enforce a single aggregate bound. The per-bidder form is the most direct interpretation of an "individual" stability promise and avoids a scenario in which a small subset of bidders faces large shading incentives while the average remains acceptable. The downside is statistical: $\widehat{G}_{i,t}$ can be noisy bidder-by-bidder, especially when some bidders participate infrequently, and the resulting $\lambda_{i,t}$ may overreact.

Two aggregations are common. First, a max-type constraint $\max_i G_{i,t} \le \varepsilon$ can be implemented by a smooth approximation (e.g. log-sum-exp) or by maintaining a single dual variable driven by the worst observed bidder:

$$\lambda_{t+1} = \big[\lambda_t + \eta_\lambda(\max_i \widehat{G}_{i,t}(\theta_t) - \varepsilon)\big]_+.$$

Second, a weighted average $\sum_i w_i G_{i,t} \le \varepsilon$ can be matched to policy objectives (e.g. weights based on spend, participation, or protected classes), at the cost of weakening individual guarantees. In regulated settings we typically recommend per-bidder constraints whenever participation is sufficiently dense to support stable estimation, and otherwise a hybrid: enforce a max constraint over coarse bidder segments (by size or vertical) while logging bidder-level metrics for audit.

Smoothing and two-timescale control. Because $\widehat{\text{I-DIC}}$ is a difference of two counterfactual values, it inherits variance from both the numerator and the continuation-value approximation. Feeding the raw per-round estimate into a controller can yield "chatter" in θ and oscillations in λ. We therefore separate *measurement* from *actuation*: we maintain an exponential moving average

$$\widetilde{G}_{i,t} = (1-\rho)\widetilde{G}_{i,t-1} + \rho\,\widehat{G}_{i,t}, \qquad \rho \in (0,1],$$

and drive the dual update with $\widetilde{G}_{i,t}$ rather than $\widehat{G}_{i,t}$. A complementary stabilization is a two-timescale step-size choice $\eta_\lambda \gg \eta_\theta$: the dual reacts quickly to emerging violations (raising penalties), while the primal moves cautiously to avoid overshooting. In practice we also cap λ above by $\lambda_{\max}$ to prevent extreme penalties from forcing θ to the boundary of Θ due to a brief burst of measurement noise.

Safety filters and rollback mechanisms. Primal–dual updates are asymptotic guarantees; a platform operator typically cares about *ex post* safety in each deployment window. We therefore implement a "safety filter" that sits between the computed update and the production mechanism. The simplest filter is a backtracking line-search on the step size: propose $\theta_{t+1}^{\text{prop}}$ from the update above, re-evaluate a fast proxy of $\widetilde{G}_{i,t}$ (or a conservative upper confidence bound), and shrink the step until the proxy is below ε (or below

$\varepsilon - \delta$ for a slack $\delta > 0$). When fast re-evaluation is unavailable, we use a conservative rule: if $\widetilde{G}_{i,t} > \varepsilon$, then (i) freeze the primal update ($\theta_{t+1} = \theta_t$) and (ii) increase λ until the constraint pressure is sufficiently large that the next non-frozen step is dominated by the penalty term.

A second safety instrument is rollback to a certified-safe baseline θ^{safe}. Many platforms have a historically stable configuration (e.g. a fixed reserve policy) that is known to exhibit low manipulability. We keep θ^{safe} as an anchor and impose a trust region $\|\theta_t - \theta^{\text{safe}}\| \le r$. If the monitored metric breaches a hard threshold $\varepsilon_{\text{hard}} > \varepsilon$ (suggesting a genuine regime break rather than noise), we reset $\theta_{t+1} \leftarrow \theta^{\text{safe}}$ and restart the dual variables. This introduces conservatism, but it aligns with operational risk management: short-lived revenue improvements are not worth large incentive shocks.

Step sizes, batching, and delayed feedback. In high-throughput markets, updating θ every round is unnecessary and can amplify noise. We commonly batch updates over windows $t \in \{kB + 1, \dots, (k+1)B\}$ and apply a single update using averaged estimators $\overline{R}_k, \overline{G}_{i,k}$. Batching reduces variance and accommodates delayed feedback (e.g. when payments or conversions are observed later). When the continuation-value approximation $\widehat{V}$ is itself learned online, we recommend staggering updates: update $\widehat{V}$ at a faster cadence, and update θ more slowly once the value estimates stabilize, to avoid coupling two drifting estimators.

Implementation with non-differentiable mechanisms. Some components of $\mathcal{M}_\theta$ are discrete (rank thresholds, tie-breaking, eligibility rules), so $\nabla_\theta \widehat{R}_t$ and $\nabla_\theta \widehat{G}_{i,t}$ may not exist in the classical sense. In those cases we use (i) differentiable relaxations (softmax ranks, smoothed reserves), (ii) straight-through estimators, or (iii) bandit-style gradient estimates (random perturbations of θ and regression of outcomes on perturbations). The key requirement for the control logic is not perfect differentiability, but a directionally informative update that, on average, moves revenue up while moving the incentive metric down when it is above budget.

What the controller does *not* guarantee. Finally, we emphasize a limitation that informs the theory that follows. The controller enforces a *local* sensitivity cap with respect to small multiplicative shading, and it does so through noisy estimates. This does not rule out profitable large deviations or sophisticated dynamic policies in a fully adversarial sense. What it does provide is a disciplined way to keep the objective locally flat around truthful bidding and to prevent learning-driven changes in θ from inadvertently creating steep incentive gradients. In the next section we formalize how such local flatness, when maintained uniformly over time, translates into a bound on the global discounted advantage of bid shading.

7 Theory I (stability): from local I-DIC caps to global manipulation bounds

Our monitoring signal I-DIC$_{i,t}$ is intentionally *local*: it measures the directional derivative of bidder i's discounted objective at the truthful bid under a small multiplicative shading. The central theoretical question is whether such local flatness can be promoted into a *global* guarantee on the discounted advantage of any admissible bid-shading policy. The answer is yes, provided we restrict attention to deviations that remain in a neighborhood of truth-telling and we assume mild regularity so that derivatives control finite changes.

Deviation class and the relevant value function. Fix a bidder i and a round t. We consider deviations of the form

$$b_{i,t} \;=\; s_{i,t}(h_{t-1})\, v_{i,t}, \qquad s_{i,t}(h_{t-1}) \in [1-\bar{\alpha},\, 1+\bar{\alpha}],$$

with $\bar{\alpha} \in (0,1)$ small. This class captures the most operationally common manipulations (systematic shading or overbidding relative to value) while remaining interpretable as a local perturbation around truthful bidding.

To connect a one-step derivative to an intertemporal payoff, it is convenient to define a *one-shot* discounted objective at time t that already includes the continuation value induced by the deviation:

$$W_{i,t}(s; h_{t-1}) \;:=\; \mathbb{E}\big[\hat{u}_{i,t}(s v_{i,t}; h_{t-1}, v_{i,t}) \;+\; \bar{U}_{i,t}(h_t; \theta, \text{truth thereafter}) \,\big|\, h_{t-1}\big],$$

where $\bar{U}_{i,t}$ is the expected future discounted utility from $t+1$ onward under a specified reference continuation rule (e.g., truthful bidding in future periods). The choice of continuation rule matters for interpretation: I-DIC is a sensitivity of a *particular* dynamic objective, and our bound is correspondingly with respect to the same objective. This is appropriate for auditing and control, where we must commit to a reference behavior for counterfactual evaluation.

A stability-to-gain bound. The core implication of I-DIC$_{i,t} \le \varepsilon$ is that the derivative of $W_{i,t}$ with respect to shading is small at $s = 1$, once we normalize by truthful expected welfare $\mathbb{E}[v_{i,t} a_{i,t}(v_{i,t})]$. To translate this into a bound for *finite* deviations $s \neq 1$, we assume that $W_{i,t}(\cdot; h_{t-1})$ is regular enough that its slope does not change arbitrarily fast within $[1-\bar{\alpha}, 1+\bar{\alpha}]$. One convenient sufficient condition is that $W_{i,t}$ is differentiable on this interval and that its derivative is bounded there by the monitored budget:

$$\left| \frac{\partial}{\partial s} W_{i,t}(s; h_{t-1}) \right| \;\le\; \varepsilon \cdot \mathbb{E}[v_{i,t} a_{i,t}(v_{i,t}) \mid h_{t-1}] \qquad \forall s \in [1-\bar{\alpha}, 1+\bar{\alpha}]. \quad (1)$$

Condition (1) can be read in two ways. First, it is exactly what we obtain if we monitor I-DIC not only at $s = 1$ but also at the two perturbed bids $(1 \pm \alpha)v$ and impose a conservative bound over the neighborhood (which is feasible in high-throughput environments). Second, it can be implied by monitoring at $s = 1$ together with a Lipschitz condition on the derivative in s (so that small neighborhoods inherit small slopes up to a second-order slack).

Under (1), the gain from any admissible shading factor $s_{i,t}$ is controlled by integrating the slope:

$$W_{i,t}(s_{i,t}; h_{t-1}) - W_{i,t}(1; h_{t-1}) \;=\; \int_1^{s_{i,t}} \frac{\partial}{\partial s} W_{i,t}(s; h_{t-1})\, ds \;\le\; \varepsilon \cdot \mathbb{E}[v_{i,t} a_{i,t}(v_{i,t}) \mid h_{t-1}] \cdot |s_{i,t} - 1|.$$

Since $|s_{i,t} - 1| \le \bar{\alpha}$, we obtain a per-period bound

$$W_{i,t}(s_{i,t}; h_{t-1}) - W_{i,t}(1; h_{t-1}) \;\le\; \varepsilon\, \bar{\alpha} \cdot \mathbb{E}[v_{i,t} a_{i,t}(v_{i,t}) \mid h_{t-1}]\,. \tag{2}$$

Summing (2) over t, and using geometric discounting, yields the global discounted gain bound

$$\mathbb{E}[U_i(\pi_i) - U_i(\text{truth})] \;\le\; \frac{\varepsilon\, \bar{\alpha}}{1 - \gamma} \cdot \sup_t \mathbb{E}[v_{i,t} a_{i,t}(v_{i,t})]\,. \tag{3}$$

If we want a bound that depends only on primitives and avoids a $\sup_t$, we can impose a uniform upper bound $\mathbb{E}[v_{i,t} a_{i,t}(v_{i,t})] \le 1$, and/or a lower bound $\mathbb{E}[v_{i,t} a_{i,t}(v_{i,t})] \ge \underline{w} > 0$ (which is already implicit in the normalization used to define I-DIC). With $\underline{w}$ available, (3) can be stated in the normalized form highlighted earlier: the maximum advantage scales linearly in ε and is amplified by at most $1/(1 - \gamma)$.

Where Lipschitzness enters. The derivative-band condition (1) is a clean route to (3), but it is not the only one. Under (H1), per-round utility is L-Lipschitz in the bid, which implies an a priori absolute bound

$$\left|\hat{u}_{i,t}(s v_{i,t}) - \hat{u}_{i,t}(v_{i,t})\right| \;\le\; L\,|s - 1|\, v_{i,t} \;\le\; L\,\bar{\alpha}.$$

This does not by itself guarantee that shading is unprofitable, but it controls how much any single period can contribute to a deviation. When combined with a monitored small I-DIC (which forces the *first-order* term to be small), Lipschitzness controls the residual *second-order* accumulation over time, yielding constants of the form $C = C(L, \bar{\alpha}, \underline{w})$ in the global bound.

Multiple bidders. Because $\text{I-DIC}_{i,t}$ is defined bidder-by-bidder, the stability guarantee extends pointwise: if the platform enforces $\text{I-DIC}_{i,t} \le \varepsilon$ for every i and t, then each bidder's unilateral shading gain is bounded as in (3). Importantly, the mechanism may couple bidders through allocation and

pricing, but the bound remains individual: it evaluates the best-response advantage of bidder i holding the environment fixed (including the policy used to define $\bar{U}_{i,t}$). Thus, the guarantee is not that the joint profile is an equilibrium, but that *no single bidder* has a large local incentive to shade within the prescribed class. This is precisely the notion needed for an auditor's manipulability budget and for a controller that seeks to prevent sharp incentive gradients from emerging as θ drifts.

Heterogeneous discounting. If bidders have heterogeneous discount factors γ_i, the only change is the horizon amplification term. Repeating the summation with bidder-specific discounting gives

$$\mathbb{E}[U_i(\pi_i) - U_i(\text{truth})] \;\le\; \frac{\varepsilon\,\bar{\alpha}}{1-\gamma_i} \cdot \sup_t \mathbb{E}[v_{i,t} a_{i,t}(v_{i,t})]\,.$$

This comparative static is operationally relevant: in markets with sophisticated, patient bidders (large γ_i), even small per-period incentive slopes can cumulate, so the same ε corresponds to a weaker global guarantee. Conversely, for myopic bidders (small γ_i), the platform can tolerate a larger local slope without materially increasing the long-run manipulation bound.

What this guarantee does and does not say. We should be explicit about scope. The bound controls discounted gains from *small, multiplicative* shading policies that remain near truthful bids; it does not exclude profitable large deviations, nor does it certify equilibrium play under arbitrary dynamic strategies. Its value is instead as a stability certificate: by keeping $W_{i,t}$ locally flat around truth-telling uniformly over time, we prevent learning-driven updates of θ from creating systematically exploitable gradients. The next step is to ask whether the same monitoring-and-control apparatus can deliver *platform-side* performance guarantees—namely, revenue regret bounds for the online controller relative to the best fixed feasible mechanism under tractable surrogate assumptions.

8 Theory II (online performance): regret of incentive-aware control

We now turn from the bidder-side guarantee to the platform-side question: if we *actively* enforce I-DIC $\le \varepsilon$ while learning or adapting θ, how much revenue do we lose relative to the best fixed feasible mechanism we could have run in hindsight? This is the relevant benchmark for practice. A controller that stabilizes incentives but destroys revenue is not a credible policy tool; conversely, a controller that approximately matches the best feasible static design while maintaining the manipulability budget gives a principled "no-regrets for auditing" interpretation.

Online problem and benchmark. At each round t, after observing (x_t, h_{t-1}), the platform selects parameters $\theta_t \in \Theta$ (a compact convex set) and runs $\mathcal{M}_{\theta_t}$. Let $R_t(\theta_t)$ denote the realized platform revenue in that round (e.g., $\sum_i p_{i,t}$), and let $G_t(\theta_t)$ denote an aggregate incentive-sensitivity signal (e.g., $\max_i \text{I-DIC}_{i,t}(\theta_t)$ or $\frac{1}{n}\sum_i \text{I-DIC}_{i,t}(\theta_t)$). Since I-DIC is typically estimated with perturbations, we work with estimators $\widehat{R}_t(\theta_t)$ and $\widehat{G}_t(\theta_t)$, measurable with respect to the platform's information at time t.

The natural comparator is the best *fixed* parameter θ that respects the incentive budget:

$$\theta^\star \in \arg\max_{\theta\in\Theta} \sum_{t=1}^{T} \mathbb{E}[R_t(\theta)] \quad \text{s.t.} \quad \mathbb{E}[G_t(\theta)] \le \varepsilon \;\; \forall t,$$

or, in a stationary formulation, $\mathbb{E}[G(\theta)] \le \varepsilon$ under the induced data-generating process. We emphasize the interpretation: we are not comparing to the globally optimal *unconstrained* revenue-maximizer (which may be highly manipulable), but to the best mechanism that an auditor would deem acceptable.

Convex surrogate assumption. To obtain transparent regret bounds, we adopt a standard surrogate view: there exist convex functions $r_t(\theta)$ and $g_t(\theta)$ such that (i) maximizing $\sum_t r_t(\theta)$ is aligned with maximizing revenue (e.g., r_t is a concave lower bound or a convex loss $-r_t$ is minimized), and (ii) $g_t(\theta)$ upper bounds the incentive signal (or a smooth approximation thereof). Formally, we assume that $-r_t(\theta)$ and $g_t(\theta)$ are convex on Θ with bounded gradients,

$$\|\nabla_\theta r_t(\theta)\| \le G_R, \qquad \|\nabla_\theta g_t(\theta)\| \le G_G,$$

and that the stochastic estimators satisfy a bounded-bias condition,

$$\mathbb{E}\Big[\nabla_\theta \widehat{R}_t(\theta) \mid h_{t-1}\Big] = \nabla_\theta r_t(\theta) + \delta_t^R(\theta), \qquad \mathbb{E}\Big[\nabla_\theta \widehat{G}_t(\theta) \mid h_{t-1}\Big] = \nabla_\theta g_t(\theta) + \delta_t^G(\theta),$$

with $\|\delta_t^R(\theta)\| \le \Delta_R$ and $\|\delta_t^G(\theta)\| \le \Delta_G$. In particular, when $\widehat{G}_t$ is computed by symmetric finite differences with step size α, the smoothness assumptions underlying Proposition 1 yield $\Delta_G = O(\alpha^2)$ (at the cost of variance that grows as α shrinks).

Primal–dual controller and its guarantee. We consider the online Lagrangian

$$\mathcal{L}_t(\theta, \lambda) = -r_t(\theta) + \lambda\big(g_t(\theta) - \varepsilon\big), \qquad \lambda \ge 0,$$

and implement projected stochastic gradient steps,

$$\theta_{t+1} = \Pi_\Theta\Big(\theta_t + \eta\, \nabla_\theta \widehat{R}_t(\theta_t) - \eta\, \lambda_t\, \nabla_\theta \widehat{G}_t(\theta_t)\Big), \qquad \lambda_{t+1} = \Big[\lambda_t + \eta_\lambda\big(\widehat{G}_t(\theta_t) - \varepsilon\big)\Big]_+. \tag{4}$$

The economic meaning is direct: λ_t is an endogenous "shadow price" of manipulability. When the estimated incentive sensitivity exceeds the budget, λ_t rises and the next parameter update places more weight on reducing G_t; when the system is comfortably stable, λ_t relaxes and the update behaves more like revenue ascent.

Under the convex surrogate assumptions and boundedness of Θ, standard online primal–dual analysis implies two performance statements. First, the (expected) *revenue regret* against the best fixed feasible θ is sublinear:

$$\mathrm{Reg}_T := \max_{\theta\in\Theta:\, g_t(\theta)\le\varepsilon\ \forall t} \sum_{t=1}^{T} \mathbb{E}[r_t(\theta) - r_t(\theta_t)] \;\le\; O(\sqrt{T}) \;+\; O((\Delta_R + \Delta_G)\,T)\,,$$

with the $O(\sqrt{T})$ term driven by gradient noise and the diameter of Θ, and the linear term vanishing when estimators are (nearly) unbiased. Second, the cumulative constraint violation is also sublinear:

$$\sum_{t=1}^{T} \mathbb{E}\Big[\big(g_t(\theta_t) - \varepsilon\big)_+\Big] \;\le\; O(\sqrt{T}) \;+\; O(\Delta_G\,T),$$

so that the *time-average* violation is $O(1/\sqrt{T})$ (plus the estimator-bias floor). These bounds formalize the operational promise: enforcing incentive stability does not impose a persistent revenue tax relative to the best audited mechanism, provided our monitoring signal is accurate enough.

How estimation quality enters. The preceding display makes a point that is easy to miss in purely economic statements: incentives are only as enforceable as they are measurable. If the counterfactual model used to compute $\widehat{\text{I-DIC}}$ is misspecified, the controller may systematically underestimate g_t, keeping λ_t too low and allowing manipulability to drift upward; overestimation has the opposite effect, pushing the system toward overly conservative θ_t and sacrificing revenue. This is why we view ε as a policy instrument *coupled* to an estimator: a small ε is only meaningful when Δ_G is commensurately small, otherwise the bias floor dominates the theoretical $o(1)$ violation rate.

Nonconvex reality and what we can still claim. The convex surrogate assumptions are, candidly, not literally true in many modern mechanisms. Neural allocation rules, reserve-price networks, pacing heuristics, and quality-score transformations often create nonconvex revenue landscapes in θ, and the incentive metric itself may be nonconvex because it composes equilibrium responses, counterfactual estimators, and truncations (e.g., $[\cdot]_+$, maxima over bidders). In such settings, we should not expect global regret guarantees against the best feasible parameter in hindsight.

Nevertheless, the convex analysis remains useful in two ways. First, it motivates *designing* controllers around convexified surrogates (e.g., smooth upper bounds on $\max_i \text{I-DIC}_{i,t}$, regularization that keeps θ_t in regions with stable gradients, or linearization of g_t around the current iterate). Second, even in nonconvex problems, primal–dual stochastic gradient methods often provide meaningful *local* guarantees: convergence to approximate stationary points of a penalized objective, and bounded constraint violations in practice when the dual step sizes are tuned to the scale of measurement noise. Our stance is therefore pragmatic: we use the convex theory as a disciplined baseline (it clarifies what should scale like $\sqrt{T}$, what should scale like $1/(1-\gamma)$, and where bias enters), and we treat deviations from convexity as an empirical question rather than an article of faith.

Empirical validation targets. The theory suggests concrete diagnostics that we can and should validate experimentally: (i) revenue tracks the best feasible static baseline up to a transient that shrinks with T; (ii) the realized $\widehat{\text{I-DIC}}$ hovers near ε with occasional excursions attributable to noise, rather than drifting upward; and (iii) tightening ε produces a predictable revenue–stability tradeoff rather than unstable oscillations in λ_t. In the next section we implement precisely these checks in several repeated-market environments, including settings with distribution shift where the value of explicit incentive control is most apparent.

9 Experiments (repeated markets, distribution shift, and robustness)

Our experiments serve three purposes. First, we test whether the primal–dual controller can keep the manipulability signal near the target budget, i.e., whether realized incentive sensitivity tracks ε rather than drifting. Second, we quantify the revenue cost of enforcement relative to natural baselines (including an unconstrained learner that ignores incentives). Third, we stress the estimator: we study how performance changes as we tighten ε, increase bidder patience γ, and degrade the quality of the counterfactual model used to compute $\widehat{\text{I-DIC}}$.

Common experimental protocol. Across environments, the platform chooses a parameter θ_t each round and then runs a history-dependent mechanism $\mathcal{M}_{\theta_t}$. We evaluate (i) discounted revenue $\sum_{t=1}^{T} \gamma_p^{t-1} \sum_i p_{i,t}$, (ii) average constraint violation $\frac{1}{T}\sum_{t=1}^{T}(G_t(\theta_t) - \varepsilon)_+$, and (iii) ex post strategic gain of representative bidders from multiplicative shading. For the latter, we simulate deviations of the form $b_{i,t} = s_{i,t} v_{i,t}$ with $s_{i,t} \in [1-\bar{\alpha}, 1+\bar{\alpha}]$ and compare discounted utility to truthful bidding, holding the platform policy

fixed; this aligns with the local deviation class used by I-DIC while still permitting history dependence in $s_{i,t}(h_{t-1})$.

To isolate the role of measurement, we consider two monitoring regimes. In the *oracle* regime, we compute $G_t(\theta_t)$ using the simulator directly (or analytic derivatives when available), so that estimator error is negligible. In the *noisy* regime, we compute $\widehat{G}_t$ via symmetric finite differences with perturbation size α and add controlled noise/bias to reflect model misspecification (e.g., a propensity model trained on a limited window). This lets us empirically instantiate the bias floor discussed earlier.

Baselines. We compare five platform policies: (i) *No-control* (revenue-only): a stochastic gradient update on θ using $\widehat{R}_t$ with no incentive term (equivalently, $\lambda_t \equiv 0$); (ii) *Fixed feasible*: the best fixed θ found by offline search subject to $G(\theta) \leq \varepsilon$ under the training distribution (a stringent but stationary benchmark); (iii) *Static regret minimization*: an online learner that minimizes a surrogate revenue loss but does not maintain a dual variable (so it adapts, yet treats stability as a post hoc diagnostic); (iv) *Penalty with fixed price*: a single tuned $\bar{\lambda}$ and updates on $-r_t(\theta) + \bar{\lambda} g_t(\theta)$ (capturing the common practice of ad hoc regularization); (v) *Primal–dual control* (ours): the adaptive λ_t policy.

9.1 Ad-auction-like repeated setting (quality scores and pacing)

We begin with a stylized ad auction in which each round corresponds to a query/impression with context x_t (user features and slot effects). Each bidder i draws a private per-click value $v_{i,t} \in [0, 1]$ and a quality score $q_{i,t} \in [q_{\min}, q_{\max}]$ that depends on x_t and bidder-specific relevance. Allocation is determined by a score $S_{i,t}(\theta) = \phi_\theta(q_{i,t})\, b_{i,t}$, where $\phi_\theta(\cdot)$ is a monotone transformation (e.g., $\phi_\theta(q) = q^\theta$ or a clipped linear map), and payments follow a generalized second-price rule with an optional reserve $r(\theta)$. We also include a pacing component: each bidder has a budget, and the platform applies a multiplicative pacing multiplier that depends on observed spend, creating the kind of history dependence that makes dynamic incentives salient.

Distribution shift. Midway through the horizon we perturb the environment: either the distribution of $q_{i,t}$ shifts (e.g., a change in user mix), or a subset of bidders experiences a mean increase in values (e.g., seasonal demand). This shift is chosen so that a revenue-only learner benefits from increasing the aggressiveness of ϕ_θ or raising reserves, but doing so tends to steepen the utility slope around truthful bidding.

Findings. In the oracle monitoring regime, primal–dual control keeps $G_t(\theta_t)$ tightly concentrated near ε, while the no-control policy reliably violates the budget after the shift. Revenue under control tracks the best fixed feasible θ up to a transient: before the shift, the controller behaves similarly

to the revenue-only learner; after the shift, λ_t increases and θ_t moves toward a less manipulative region (typically lowering the effective rank curvature and/or relaxing reserves). Importantly, the bidder-side deviations we simulate yield substantially lower realized utility gains under control, and the gap widens with higher γ, consistent with the notion that dynamic channels amplify small per-period slopes.

9.2 Cloud resource allocation (capacity constraints and throttling)

Our second environment is a repeated allocation of divisible resources (CPU/GPU time slots) under capacity constraints. Each round t brings a context x_t describing available capacity and job mix. Bidder i has value $v_{i,t}$ for receiving a unit of resource; allocation is proportional to bids through a parameterized throttling rule $a_{i,t} = f_\theta(b_t, x_t, h_{t-1})$ that can mimic weighted proportional allocation with congestion pricing. Payments are per-unit at a price determined by a market-clearing or posted-price-like function of aggregate demand, with θ controlling the aggressiveness of the congestion response.

Distribution shift. We introduce bursts: periods of heavy demand followed by slack capacity. These bursts make history dependence operationally relevant because the platform's throttling and pricing rules respond to past congestion.

Findings. The main qualitative pattern persists: no-control learns high-gain parameters during bursts, which raises both revenue and the measured incentive slope. Primal–dual control dampens this behavior by endogenously raising λ_t during high-congestion episodes, reducing the marginal benefit of shading. Compared to a fixed penalty $\bar{\lambda}$, adaptive pricing of manipulability is notably less conservative in slack periods, where the same $\bar{\lambda}$ would unnecessarily suppress revenue.

9.3 Mobility charging (dynamic pricing with temporal substitution)

Our third environment models repeated allocation of charging slots in a mobility setting (e.g., EV charging). Each round corresponds to a time interval with limited chargers and time-varying base demand. Agents have values for charging now versus later (captured by $v_{i,t}$ and an exogenous continuation outside the mechanism). The platform sets a parameterized priority rule and price schedule (e.g., a reserve or surge multiplier θ_t) that affects both allocation probability and payment.

Distribution shift. We change commuting patterns: demand shifts earlier in the day, altering the scarcity profile. This shift creates a plausible real-world failure mode: a policy trained on one regime can become both revenue-inefficient and manipulable when scarcity arrives at unexpected times.

Findings. Incentive-aware control adapts in a way that is economically interpretable: the shadow price λ_t rises precisely when scarcity makes prices more sensitive to bids. In contrast, static regret minimization (without the dual variable) tends to chase revenue spikes and, as a byproduct, increases measured manipulability. In this environment, the coupling between γ and manipulation is especially clear: when agents are more patient (larger γ), intertemporal substitution makes dynamic incentives more consequential, and the gap between controlled and uncontrolled strategic gain widens.

9.4 Sensitivity to ε, γ, and estimator noise

Finally, we sweep key parameters to map the practical tradeoffs. Tightening ε yields a smooth revenue–stability frontier: revenue declines monotonically while realized strategic gain declines sharply at first and then levels off, reflecting diminishing returns once the utility slope is close to flat. Increasing γ does not materially change the platform's ability to *measure* G_t, but it increases the realized utility gains under no-control, making the welfare case for enforcement stronger in patient markets.

Estimator degradation matters in the expected way. Added variance in $\widehat{G}_t$ produces occasional constraint overshoots but does not systematically break control when η_λ is tuned conservatively. In contrast, persistent downward bias (underestimating manipulability) yields chronic underenforcement: λ_t remains too small and violations accumulate, mirroring the bias-floor term in the theory. This reinforces the operational message: choosing ε is inseparable from choosing (and validating) the counterfactual model used to audit incentives.

Taken together, these experiments indicate that explicit incentive monitoring can be integrated into online adaptation without imposing a persistent revenue tax relative to audited benchmarks, and that the main failure modes are measurement failures rather than optimization failures. The next section discusses what our deviation class leaves out, why Lipschitzness and counterfactual validity are the real structural assumptions, and how these choices relate to classical dynamic DSIC.

10 Discussion and limitations

Our approach is motivated by a pragmatic observation: in many repeated markets the platform cannot (or will not) commit to a fully dynamic mechanism designed ex ante for dominant-strategy truthfulness, yet it can often *measure* when the current policy makes utility steep in the direction of simple shading. We therefore treat incentives as an auditable *sensitivity constraint* that can be monitored online and enforced with a controller. This framing illuminates a tradeoff that is familiar in practice but rarely formalized in

deployment terms: a platform can buy robustness to manipulation by flattening local utility gradients, but doing so restricts the set of revenue-relevant parameter moves it can make under distribution shift.

Admissible deviation class. A central modeling choice is the deviation class captured by I-DIC: multiplicative shading around truthful bidding, $b_{i,t} = s_{i,t}(h_{t-1})v_{i,t}$ with $s_{i,t}$ near 1. We view this as a disciplined compromise between behavioral realism and statistical feasibility. It is behaviorally plausible because many bidders implement bidding rules that scale a value proxy (or a learned value model) by a single aggressiveness knob; it is also the class for which local derivatives are meaningful and can be estimated with symmetric perturbations. Technically, the local nature matters: bounding a directional derivative at $s = 1$ provides a first-order certificate that "small" shading does not pay, and the Lipschitz/derivative arguments then convert this certificate into a bound on total discounted gain for shading policies constrained to $[1 - \bar{\alpha}, 1 + \bar{\alpha}]$.

The limitation is equally clear. Our metric does not preclude profitable *nonlocal* deviations: large bid jumps, switching between bidding modes, or policies that depend on private state in a way not representable by a multiplicative factor. Nor does it directly address deviations that exploit discrete rule changes (e.g., crossing a reserve or a throttling threshold), nor does it capture strategic delay, participation decisions, or multi-account behavior. In that sense, I-DIC should be interpreted as a *local manipulability budget*: it constrains the marginal return to shading in the neighborhood where many real bidding systems operate, but it is not a full equilibrium concept. A natural extension is to define a family of sensitivity constraints along multiple directions (e.g., additive perturbations, or perturbations in a low-dimensional bid-policy basis) and to enforce a worst-case bound over that family; doing so would broaden coverage but raises monitoring cost and variance.

Dependence on Lipschitzness and smoothness. The stability bound relies on a regularity condition that is easy to state but not always innocent: per-round utility must be L-Lipschitz in own bid (and differentiable at the reference bid). This is the bridge from "small derivative" to "small gain," especially when we sum gains over time with discounting. Economically, Lipschitzness rules out knife-edge mechanisms where an infinitesimal bid change triggers a large payment or allocation discontinuity; statistically, it controls the error induced by finite differences and by imperfect counterfactual models.

Many marketplace mechanisms do have approximate Lipschitz structure after accounting for randomization, tie-breaking noise, or smoothing (e.g., stochastic assignment, soft reserves, or continuous throttling). However,

pure rank-by-bid allocation with deterministic tie-breaking, hard eligibility constraints, and step-function pacing can violate differentiability and create local cliffs. In such cases, I-DIC may be unstable to estimate (high variance as $\alpha \to 0$) and potentially misleading (a derivative may not exist, while discrete deviations may be profitable). One practical implication is design-oriented: if a platform wants incentive auditing to be meaningful, it may need to *engineer* smoothness into the mechanism (e.g., randomized smoothing or continuous relaxations). This is not merely a mathematical convenience; it is a governance choice about whether the market should be sensitive to infinitesimal strategic moves.

Counterfactual estimation and identification. A second structural dependence is on the ability to compute $\widehat{\text{I-DIC}}$, which is inherently counterfactual: we must evaluate utility under perturbed bids while holding $(x_t, h_{t-1}, v_{i,t})$ fixed. In an oracle simulator this is straightforward; in live markets it becomes an off-policy estimation problem. The key identification requirement is that the platform can model how allocations and payments would have changed under the perturbed bid, which typically demands either (i) a known differentiable mapping $(x_t, h_{t-1}, b_t) \mapsto (a_t, p_t)$, or (ii) a learned surrogate with adequate coverage near the observed bids. Bias is the central risk: a downward-biased monitor systematically underprices manipulability and leads the controller to under-enforce the constraint. This risk is particularly acute when the mechanism itself is changing, since the data distribution over bids and contexts is endogenous to θ_t.

From an operational perspective, this suggests that incentive monitoring should be treated like safety monitoring: it requires calibration, stress tests, and explicit uncertainty accounting. One conservative modification is to enforce a high-probability constraint using confidence bounds, e.g.,

$$\widehat{G}_t(\theta_t) + \mathrm{rad}_t \le \varepsilon,$$

where rad_t reflects estimation uncertainty. This reduces false negatives (missed manipulability) at the expense of some revenue, and it makes transparent how monitoring quality interacts with the chosen budget ε.

Connections to dynamic DSIC and classical mechanism design. It is useful to situate I-DIC relative to dynamic dominant-strategy incentive compatibility. Dynamic DSIC is a *global* property: truthful reporting is optimal regardless of history and regardless of how far the agent deviates. Achieving it typically requires strong commitment and carefully structured transfers (e.g., dynamic pivot or bank-account mechanisms) that internalize future consequences. By contrast, our constraint is *local and directional*. It is closer in spirit to a first-order optimality condition: if truthful bidding is optimal in a smooth environment, then the derivative of the bidder's value

function with respect to shading should be near zero at $s = 1$. We operationalize this idea by directly estimating that derivative and regulating it.

This distinction also clarifies what our results do *not* claim. We do not claim to implement a truthful equilibrium, nor do we claim robustness to arbitrary deviations. Rather, we provide a measurable certificate that the mechanism is not "inviting" a particular family of profitable deviations, and we show that (under regularity) this certificate bounds the total discounted gain from those deviations. In this sense, I-DIC can be interpreted as an "engineering approximation" to dynamic incentive constraints, intended for systems that must adapt online and cannot solve the full dynamic mechanism design problem.

Audit implications and governance. Because I-DIC is a scalar signal that can be tracked over time, it lends itself naturally to auditing. An internal risk team (or an external auditor) can require that $\text{I-DIC}_{i,t} \leq \varepsilon$ on average, by segment, or under stress scenarios, much like reliability constraints in other safety-critical systems. The dual variable λ_t has an interpretation that can be communicated: it is the shadow price of manipulability, revealing when the platform is implicitly trading revenue for incentive stability. This transparency is valuable, but it also raises design questions: which bidder groups are monitored, how the normalization $\mathbb{E}[va(v)]$ is estimated, and whether monitoring can be done without exposing sensitive bid-response models. These are governance choices, not purely technical ones.

Open problems. Several extensions remain open. First, we would like broader deviation coverage without losing measurability: can we learn a small set of adversarial deviation directions online and enforce a max-sensitivity constraint over that set? Second, the analysis takes the platform's policy as the object of control and evaluates bidder deviations holding the policy fixed; a richer model would endogenize bidder learning dynamics and analyze coupled learning processes. Third, it is unclear how to optimally choose ε: it reflects both a normative stance (how much manipulability is acceptable) and a measurement stance (how much monitoring error we can tolerate). Finally, mechanisms often have multiple desiderata—fairness, budget balance, stability, and revenue—and incentive sensitivity may interact with these constraints in nontrivial ways. Understanding when local incentive flattening complements (or conflicts with) other forms of robustness is, in our view, the most important step toward making dynamic incentive control a standard component of repeated-market design.