

Joint Incentive Contract and Assignment Design for Mobile Networked Platforms with Endogenous Supply

Yumou Liu
liuyumou@sjtu.edu.cn
Shanghai Jiao Tong University

Zhenzhe Zheng
zhengzhenzhe@sjtu.edu.cn
Shanghai Jiao Tong University

Jiang Rong
rongjiang@xiaohongshu.com
Xiaohongshu

Ling Yan
yanling@xiaohongshu.com
Xiaohongshu

Fan Wu
fwu@cs.sjtu.edu.cn
Shanghai Jiao Tong University

Guihai Chen
gchen@cs.sjtu.edu.cn
Shanghai Jiao Tong University

Abstract

Mobile networked platforms such as crowdsensing markets, edge-computing exchanges, vehicular service systems, and sponsored-content programs must coordinate two coupled levers: an assignment rule that decides who serves each request and an incentive contract that determines how realized value is shared. In all four settings, supply is endogenous, that is, it depends on the platform's own decisions: providers remain active only when expected earnings cover their private participation cost. We study a discounted platform problem in which assignment and revenue sharing jointly shape future supply. We calibrate the tractable one-period objective using an auxiliary memoryless discounted surrogate: it maximizes a global lower bound on that surrogate objective, and near high-survival stationary operating points its retention coefficient becomes the local shadow price of discounted future welfare within the surrogate model. This yields an Equimarginal Principle requiring efficient platforms to equalize the marginal retention efficiency of assignment and contracting. Under complete information, we obtain an urgency-adjusted assignment score, a KKT condition for the contract, and a Frank-Wolfe coordinate-ascent benchmark with an $O(1/K)$ joint stationarity rate. Under hidden provider quality, we develop SE-RW-UCB, a subsidized-exploration algorithm with retention-weighted contextual assignment and a learned safe contract floor. A KuaiRec-based online recommendation benchmark shows that separated designs leave a sizable retention gap. In that benchmark, Oracle Fixed is the profit leader, whereas SE-RW-UCB attains the strongest high-type and all-provider retention at an interior contract. The service-quality differences are comparatively small, indicating that the main leverage of co-design in this calibration lies in supply preservation rather than per-request quality gains.

This work was supported in part by China NSF grants U2268204, 62322206, 62432007, 62441236, 62132018, 62025204, 62272307, 62372296, U25A6024. The opinions, findings, conclusions, and recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agencies or the government. Zhenzhe Zheng is the corresponding author.



This work is licensed under a Creative Commons Attribution 4.0 International License. *MobiHoc '26, Tokyo, Japan*

© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2975-1/2026/11
<https://doi.org/10.1145/3842721.3849687>

CCS Concepts

• **Theory of computation** → **Algorithmic mechanism design**;
• **Online learning algorithms**; • **Networks** → *Network economics*;
• **Information systems** → *Recommender systems*.

Keywords

contract design, assignment, contextual bandits, two-sided platforms, mobile crowdsensing, mobile edge computing

ACM Reference Format:

Yumou Liu, Zhenzhe Zheng, Jiang Rong, Ling Yan, Fan Wu, and Guihai Chen. 2026. Joint Incentive Contract and Assignment Design for Mobile Networked Platforms with Endogenous Supply. In *The Twenty-seventh International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing (MobiHoc '26)*, November 23-26, 2026, Tokyo, Japan. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3842721.3849687>

1 Introduction

Modern digital platforms increasingly operate by coordinating heterogeneous supply and demand under a platform-controlled monetization rule. In the original scenario that motivated this work, sponsored-content programs on platforms such as TikTok [30] and Xiaohongshu [33] form markets among users, creators, and advertisers [23, 31, 37]. The platform does not merely decide which content to expose to which users. It also determines how advertising revenue is shared with creators, and that contract directly shapes whether creators continue supplying content. This distinguishes the problem from the classic AdWords abstraction [15, 22]: the inventory is no longer fixed. It is generated by strategic suppliers who remain active only when their earnings under the platform's assignment and payment decisions cover their private participation cost [3, 12, 13].

The same economic structure appears in three mobile networked scenarios. In mobile crowdsensing, the platform assigns sensing jobs to workers whose battery drain, data use, and effort costs are private [10, 34]. In mobile edge computing, a platform routes offloading requests to heterogeneous edge nodes whose operating costs depend on energy, hardware utilization, and maintenance [29, 40]. In vehicular service systems, the platform allocates relaying, caching, or computation tasks to vehicles that incur fuel, wear, privacy, and participation costs [24]. Across all four settings, the platform controls two coupled levers: an assignment policy that decides who serves each request and a contract that determines how realized

value is shared. Supply is endogenous because workers can uninstall an app, edge nodes can shut down, vehicles can opt out, and creators can exit a monetization program. The common design problem is therefore to jointly choose assignment and incentives when provider quality is only partially observed and current decisions alter future supply.

The first challenge is the coupling of the assignment and contract levers. A fragile provider can be kept active either by routing more valuable requests to that provider or by increasing the provider's revenue share. These levers are substitutes, but they operate through different costs. Excessive reliance on assignment distorts matching quality and can waste scarce high-value requests; excessive reliance on the contract transfers revenue to the entire provider pool even when only a small subset is at risk. A correct solution must compare these two interventions in common welfare units and determine when traffic reallocation is cheaper than subsidy, and when the reverse is true.

The second challenge is hidden information under provider departure. The platform does not initially know which providers are high quality, which ones are near the participation boundary, or how much realized value they can generate in different contexts. Learning is therefore unavoidable. However, unlike standard contextual bandits, the platform cannot treat early mistakes as temporary regret only. A high-quality provider with unlucky initial observations may receive too little traffic or too little compensation, fail to cover its private cost, and leave before its type is identified. This creates a dual discovery problem: the platform must learn provider quality while simultaneously preserving enough supply for the learning process itself to continue.

The third challenge is dynamic and mathematical. Provider departure is effectively irreversible in the applications we study, so each poor decision can shrink the future feasible set. The platform's objective is therefore not a static matching problem with fixed inventory but a dynamic control problem in which present assignment and payment choices determine future supply. Moreover, once aggregate retention is modeled as a smooth function of provider-level revenue, the resulting objective is non-concave and couples all providers through the contract. Standard pricing models ignore the assignment lever, while standard scheduling and bandit models treat the supply side as fixed. Neither class of methods addresses the full co-design problem.

We address these challenges through a co-design framework that starts from the cumulative-wealth departure model and calibrates a tractable stationary objective through an auxiliary memoryless discounted surrogate. The auxiliary surrogate yields a global conservative lower bound and, near high-survival operating points, a local shadow-price calibration for the retention weight rather than an arbitrary tuning parameter. This yields an Equimarginal Principle that characterizes efficient use of assignment and contracting, a complete-information benchmark with urgency-adjusted assignment scores, a Frank–Wolfe coordinate-ascent solver with a finite-time joint stationarity guarantee, and a learning algorithm, SE-RW-UCB, that first purchases safe learning time through temporary subsidies and then switches to retention-weighted contextual assignment with a learned safe contract floor.

Our technical contributions are as follows:

- **Unified co-design model across four scenarios.** We formulate one platform model that captures the sponsored-content setting together with mobile crowdsensing, mobile edge computing, and vehicular services. The same primitives describe provider heterogeneity, assignment, contracts, and departure in all four cases.
- **Dynamic calibration and complete-information theory.** We show that an auxiliary memoryless discounted surrogate yields a global conservative lower bound and a local shadow-price calibration for the stationary objective. We then prove the Equimarginal Principle, characterize complete-information stationary points through urgency-adjusted assignment scores and a contract KKT condition, and analyze a Frank–Wolfe coordinate-ascent benchmark with an $O(1/K)$ joint stationarity rate.
- **Learning with hidden quality and irreversible exit.** We propose SE-RW-UCB, a two-phase algorithm with subsidized exploration, retention-weighted contextual assignment, and safe-floor projected contract updates. The analysis gives a safe-identification floor, a pathwise subsidy-phase survival guarantee, a post-subsidy confidence event, a Phase-2 viability condition based on a learned safe threshold, and conditional and ecosystem regret bounds.
- **Empirical validation on a KuaiRec-based online recommendation benchmark.** We evaluate the theory on a KuaiRec-based online recommendation benchmark with main, ablation, and contract-floor analyses. In that benchmark, co-design preserves more high-type and total providers than separated designs.

2 Model and Problem Formulation

2.1 System Model

The platform serves N service requests in each period. Each request belongs to one of finitely many task types $u \in \mathcal{U}$, drawn i.i.d. from a fixed arrival distribution $\rho_u \in [0, 1]$ with $\sum_{u \in \mathcal{U}} \rho_u = 1$. A task type is characterized by its context vector $\mathbf{x}_u \in \mathbb{R}^d$, which encodes the request's attributes. Let $\mathcal{I}$ denote the set of providers. Each provider $i \in \mathcal{I}$ is characterized by a feature embedding $\mathbf{x}_i \in \mathbb{R}^d$, a scalar quality parameter q_i , and a private participation cost. We assume the participation cost is linear in quality:

$$c_i = c(q_i) = c_{\text{base}} + \gamma q_i, \quad (1)$$

where c_{base} and γ are system parameters. For each task-provider pair (u, i) , we denote the effective valuation by $\alpha_{ui} > 0$. The platform controls two decision variables:

- (1) A randomized assignment policy π , defined by the probability matrix $\Pi = [\pi_{ui}]$, where $\pi_{ui} \in [0, 1]$ is the probability that a task of type u is assigned to provider i (satisfying $\sum_i \pi_{ui} = 1$ for all u).
- (2) A global revenue-sharing contract $r \in \mathcal{R} \triangleq [r_{\min}, 1]$.

The matching score of a pair is $s_{ui} \triangleq \langle \mathbf{x}_u, \mathbf{x}_i \rangle + \xi q_i$, where $\xi > 0$ is a coefficient representing the relative weight of quality in the score. The probability of a successful conversion is given by $w_{ui} \triangleq \mu(s_{ui})$, where $\mu(\cdot)$ is a strictly increasing link function. The expected gross value of a match is

$$R_{ui} \triangleq w_{ui} \alpha_{ui}. \quad (2)$$

Information structure. At each period t , the platform observes the assignment outcomes and the continuation indicators $\mathbf{1}[i \in \mathcal{I}_{t+1}]$. We assume the platform knows the cost parameters $(c_{\text{base}}, \gamma)$ but does not observe the provider-specific quality parameter q_i . It must infer q_i from noisy performance signals.¹

The expected gross revenue for provider i under policy π is

$$G_i(\pi) \triangleq N \sum_{u \in \mathcal{U}} \rho_u \pi_{ui} R_{ui}. \quad (3)$$

2.2 Retention and Welfare

We model expected provider retention as

$$S(r, \pi) = \sum_{i \in \mathcal{I}} \sigma(r G_i(\pi) - c_i), \quad (4)$$

where $\sigma(z) = (1 + e^{-\kappa z})^{-1}$ is the logistic retention function, and $\kappa > 0$ governs the sensitivity of departure decisions to net revenue.

Departure dynamics. Let $W_{i,t}$ denote provider i 's cumulative wealth up to period t , incorporating an initial buffer $W_{i,0} \geq 0$:

$$W_{i,t} = W_{i,0} + \sum_{\tau=1}^t (r_\tau G_{i,\tau} - c_i), \quad (5)$$

where r_τ and $G_{i,\tau}$ denote the contract rate and realized gross revenue at step τ , respectively. The departure indicator $D_{i,t}$ follows a state-dependent Bernoulli law:

$$\mathbb{P}(D_{i,t} = 1 \mid W_{i,t}) = 1 - \sigma(W_{i,t}). \quad (6)$$

If $D_{i,t} = 1$, provider i exits irreversibly, so $i \notin \mathcal{I}_\tau$ for all $\tau > t$. The platform does not observe the private cost c_i . The same departure law covers crowdsensing battery/data costs, edge-node operating costs, vehicular fuel-and-wear costs, and creator-side production costs. The initial buffer $W_{i,0}$ captures prior earnings, credits, or pre-existing goodwill. In Section 3, we assume a positive lower bound on $W_{i,0}$ when proving high-probability subsidy-phase survival. Without a positive buffer, the symmetry of the logistic function ($\sigma(0) = 1/2$) implies a 50% immediate exit probability even under full revenue sharing ($r = 1$).

Stationary co-design objective. The platform's one-period welfare combines immediate service value and a surrogate for long-term supply preservation:

$$\mathcal{W}(r, \pi) \triangleq \sum_{u \in \mathcal{U}} \sum_{i \in \mathcal{I}} \rho_u \pi_{ui} [(1-r)R_{ui} + \Lambda_1 s_{ui}], \quad (7)$$

where $\Lambda_1 > 0$ balances the platform's revenue share against matching quality (proxied by the score s_{ui}). The stationary co-design problem optimizes the weighted sum of immediate welfare and expected retention:

$$\max_{r \in \mathcal{R}, \pi} \mathcal{J}(r, \pi) \triangleq \mathcal{W}(r, \pi) + \Lambda_2 S(r, \pi), \quad (8)$$

where $\Lambda_2 > 0$ is the retention weight. Here $S(r, \pi)$ serves as a reduced-form, one-period proxy for long-term supply. Because the true departure process (5)–(6) is path-dependent, the objective (8) is not an exact dynamic programming formulation. Instead, Section 2.3 introduces an auxiliary memoryless surrogate only to calibrate Λ_2 ; the core learning and safety guarantees in subsequent sections are derived directly under the cumulative-wealth dynamics.

¹In the learning analysis of Section 3, we model quality specifically as the norm of a latent vector, $q_i = \|\theta_i^*\|_2$, to derive regret bounds.

2.3 Auxiliary Memoryless Surrogate for Calibrating the Retention Weight

In the auxiliary surrogate, provider i contributes one-period welfare $w_i(r, \pi) = \sum_{u \in \mathcal{U}} \rho_u \pi_{ui} [(1-r)R_{ui} + \Lambda_1 s_{ui}]$ whenever active. Let $\delta \in (0, 1)$ be the discount factor and $p_i(r, \pi) \triangleq \sigma(r G_i(\pi) - c_i)$ the one-period survival probability of provider i . The discounted value of the auxiliary memoryless surrogate is

$$\max_{r \in \mathcal{R}, \pi} \mathcal{J}_{\text{aux}}(r, \pi) \triangleq \sum_{i \in \mathcal{I}} \frac{w_i(r, \pi)}{1 - \delta p_i(r, \pi)}. \quad (9)$$

This auxiliary model preserves the monotonicity in per-period net revenue as the operational model, but is analytically distinct from the cumulative-wealth dynamics used in Sections 3 and 4.

Appendix A.1 gives a conservative global lower-bound calibration for the auxiliary surrogate. We focus in the main text on the sharper local shadow-price calibration below.

Proposition 2.1 (Local retention weight implied by the auxiliary surrogate). *Fix a compact feasible set $\mathcal{K}$ and an interior operating point $(r^\circ, \pi^\circ) \in \mathcal{K}$. Let $p_i^\circ \triangleq p_i(r^\circ, \pi^\circ)$, $w_i^\circ \triangleq w_i(r^\circ, \pi^\circ)$, and $\bar{w}^\circ \triangleq |\mathcal{I}|^{-1} \sum_i w_i^\circ$. Suppose there exist $\varepsilon, \eta \geq 0$ such that $p_i^\circ \geq 1 - \varepsilon$ and $|w_i^\circ - \bar{w}^\circ| \leq \eta$ for every provider i . Then there exist constants C_1, C_2 , depending only on derivative bounds on $\mathcal{K}$, such that*

$$\left\| \nabla \mathcal{J}_{\text{aux}}(r^\circ, \pi^\circ) - \frac{1}{1 - \delta} \nabla \mathcal{W}(r^\circ, \pi^\circ) - \frac{\delta \bar{w}^\circ}{(1 - \delta)^2} \nabla S(r^\circ, \pi^\circ) \right\| \leq C_1 \varepsilon + C_2 \eta. \quad (10)$$

Hence, up to the positive scale factor $(1 - \delta)^{-1}$, the local first-order approximation of $\mathcal{J}_{\text{aux}}$ is exactly (8) with $\Lambda_2 = \delta \bar{w}^\circ / (1 - \delta)$.

Appendix A.2 differentiates (9) for the auxiliary surrogate and separates the resulting welfare and survival gradients. The welfare term is scaled by $(1 - \delta p_i)^{-1}$, whereas the survival term is scaled by $\delta w_i (1 - \delta p_i)^{-2}$. In the high-survival, low-heterogeneity regime these coefficients become nearly common across providers, leaving a linear combination of $\nabla \mathcal{W}$ and ∇S . We use this local calibration in the main development and relegate the global conservative bound to Appendix A.1. All safety guarantees in Section 3 are proved directly under the cumulative-wealth departure law.

3 Co-Design Benchmark and Learning

3.1 Why Co-Design Is Necessary

Assignment and contracting act as substitute retention levers. Optimizing them separately is inefficient because the platform can always reallocate retention effort toward the margin with the higher marginal payoff. The joint objective (8) unifies these levers in common welfare units, yielding a precise optimality condition.

Proposition 3.1 (Equimarginal Principle). *At any interior stationary point (r^*, π^*) with $r^* \in (r_{\min}, 1)$, the marginal rate of substitution between retention and immediate welfare is equalized across all active decision margins:*

$$\frac{\partial_r S}{\partial_r \mathcal{W}} = \frac{\partial_{\pi_{ui}} S - \partial_{\pi_{uj}} S}{\partial_{\pi_{ui}} \mathcal{W} - \partial_{\pi_{uj}} \mathcal{W}} = -\frac{1}{\Lambda_2}, \quad (11)$$

for any task type u and providers i, j with positive assignment mass.

PROOF SKETCH. The first-order conditions for (8) are $\nabla_\pi \mathcal{W} + \Lambda_2 \nabla_\pi S = \lambda$ (assignment simplex) and $\partial_r \mathcal{W} + \Lambda_2 \partial_r S = 0$ (interior contract). Rearranging the contract condition yields $\Lambda_2 = -(\partial_r \mathcal{W})/(\partial_r S)$. Substituting into the assignment condition and taking differences across i, j for fixed u gives (11). $\square$

This principle formalizes the co-design imperative: a platform that fixes the contract overuses assignment distortion to retain fragile providers, while one that fixes assignment over-subsidizes the entire pool. For example, a crowdsensing worker near the participation boundary can be preserved either by routing higher-value tasks (reallocating π) or raising the global revenue share r . Optimal operation requires these interventions to yield identical marginal retention gains per unit of welfare sacrificed. The remainder of this section derives the explicit stationary scores and establishes a joint stationarity guarantee for the alternating solver.

3.2 Complete-Information Benchmark

Fix a contract $r \in \mathcal{R}$. Because the retention term couples assignments through provider-level revenue, the assignment rule must internalize a provider-specific shadow price.

Definition 3.2 (Provider urgency). For provider $i \in \mathcal{I}$, define

$$\mu_i(\pi) \triangleq \Lambda_2 \sigma'(r G_i(\pi) - c_i). \quad (12)$$

The urgency weight $\mu_i(\pi)$ peaks near the participation boundary $r G_i(\pi) \approx c_i$.

Proposition 3.3 (Urgency-adjusted assignment score). Fix $r \in \mathcal{R}$ and let π^* be a stationary assignment for (8). For every task type u , any provider with positive assignment mass under π^* maximizes

$$Z_{ui}^* = s_{ui} + \frac{R_{ui}}{\Lambda_1} [(1-r) + Nr \mu_i(\pi^*)]. \quad (13)$$

Proposition 3.3 is proved in Appendix A.3. Relative to a purely myopic ranking rule, the term $Nr \mu_i(\pi^*) R_{ui}$ prices the future supply effect of sending more revenue to provider i . The contract satisfies a scalar projected KKT condition, given in Appendix A.3.

3.3 Iterative Co-Design Solver

We solve the complete-information benchmark via block coordinate ascent: Frank–Wolfe (FW) updates for the assignment polytope Π , interleaved with projected gradient ascent for the contract. Initialize $\pi_{ui}^{(0)} = 1/|\mathcal{I}|$ and $r^{(0)} = 0.5$. Each outer iteration $t = 0, 1, \dots, K-1$ proceeds as follows:

- (1) **Urgency update:** Compute $\mu_i^{(t)} = \Lambda_2 \sigma'(r^{(t)} G_i^{(t)} - c_i)$ for all i .
- (2) **FW assignment step:** Form score-maximizing extreme points $i^*(u) = \arg \max_i Z_{ui}^*(r^{(t)}, \pi^{(t)})$ from (13), take an FW step $\pi^{(t+1)} = (1 - \gamma_t) \pi^{(t)} + \gamma_t \pi^{\text{ext}}$, and update induced revenues $G_i^{(t+1)}$.
- (3) **Projected contract step:** Apply

$$r^{(t+1)} = \text{Proj}_{\mathcal{R}} \left(r^{(t)} + \eta \nabla_r \right), \quad (14)$$

$$\text{where } \nabla_r^{(t)} = -\frac{1}{N} \sum_i G_i^{(t+1)} + \sum_i G_i^{(t+1)} \mu_i^{(t)}.$$

where π^{ext} is the extreme point that assigns each task type u to $i^*(u)$, and $\gamma_t = \min\{1, \text{FWgap}(r^{(t)}, \pi^{(t)})/C_\Pi\}$ with FWgap and C_Π defined in Proposition 3.4 below. Because (8) is non-concave and

the variables are coupled, we target first-order joint stationarity. Let L_r denote the Lipschitz constant of $\nabla_r \mathcal{J}$, and define the FW curvature constant $C_\Pi \triangleq 2|\mathcal{U}|L_\pi$ with $L_\pi = \Lambda_2 \kappa^2 N^2 R_{\max}^2/4$.

Proposition 3.4 (Joint finite-time stationarity). Let

$$\Pi \triangleq \left\{ \pi \in [0, 1]^{|\mathcal{U}| \times |\mathcal{I}|} : \sum_{i \in \mathcal{I}} \pi_{ui} = 1 \text{ for all } u \in \mathcal{U} \right\}$$

be the assignment polytope, and define $R_{\max} = \max_{u,i} R_{ui}$, $L_\pi = \Lambda_2 \kappa^2 N^2 R_{\max}^2/4$, $C_\Pi = 2|\mathcal{U}|L_\pi$, $\tilde{G}_{\max} = \max_{i,\pi} G_i(\pi)$, and $L_r = \Lambda_2 \kappa^2 |\mathcal{I}| \tilde{G}_{\max}^2/4$. Then, for every fixed $r \in \mathcal{R}$, the map $\pi \mapsto \mathcal{J}(r, \pi)$ has L_π -Lipschitz gradient on Π . Let $0 < \eta \leq 1/L_r$,

$$\mathcal{G}_\eta(r, \pi) \triangleq \frac{1}{\eta} (\text{Proj}_{\mathcal{R}}(r + \eta \nabla_r \mathcal{J}(r, \pi)) - r),$$

$$\text{FWgap}(r, \pi) \triangleq \max_{\tilde{\pi} \in \Pi} \langle \nabla_\pi \mathcal{J}(r, \pi), \tilde{\pi} - \pi \rangle,$$

and assume $K > 4(\mathcal{J}^* - \mathcal{J}^{(0)})/C_\Pi$, where $\mathcal{J}^{(0)} = \mathcal{J}(r^{(0)}, \pi^{(0)})$ and $\mathcal{J}^* = \sup_{r \in \mathcal{R}, \pi \in \Pi} \mathcal{J}(r, \pi)$. After K outer iterations of the alternating benchmark above,

$$\begin{aligned} \min_{0 \leq t < K} \left\| \mathcal{G}_\eta(r^{(t)}, \pi^{(t+1)}) \right\|^2 &= O(1/K), \\ \min_{0 \leq t < K} \text{FWgap}(r^{(t)}, \pi^{(t)})^2 &= O(1/K). \end{aligned}$$

Hence ϵ -joint stationarity is reached in $O(1/\epsilon)$ outer iterations.

Appendix A.4 proves Proposition 3.4. The resulting solver is the benchmark used to interpret the learning algorithm.

3.4 Learning Under Hidden Provider Quality

Learning model. For the bandit analysis, we work with the linearized reward approximation $R_{ui} \approx \langle \theta_i^*, \mathbf{x}_u \rangle$. Under this approximation, the platform observes

$$y_{ui} = \langle \theta_i^*, \mathbf{x}_u \rangle + \eta_{ui}, \quad (15)$$

where θ_i^* is an unknown provider-specific parameter and η_{ui} is zero-mean σ -sub-Gaussian noise. The platform must therefore learn provider quality while preventing viable providers from exiting during the discovery process.

A two-type specialization makes the difficulty transparent. Consider a new provider with hidden type $\theta \in \{\theta_L, \theta_H\}$, where $\theta_H = \bar{c} + \Delta$, $\theta_L = \bar{c} - \Delta$, and $\mathbb{P}(\theta = \theta_H) = p$. Under the adverse regime $\bar{\theta} < \bar{c}$, any static mechanism faces a dilemma: a conservative contract excludes all entrants, including high types, while a pool-wide subsidy loses money on every low-type provider over the full horizon. A standard two-point testing argument then shows that safe discrimination between high and low types requires

$$T_{\text{sub}} = \Omega \left(\frac{\sigma^2}{\Delta^2} \log \frac{1}{\delta} \right) \quad (16)$$

subsidy-phase observations; Appendix A.5 gives the details. Together these facts rule out purely static screening and motivate temporary full subsidy.

The SE-RW-UCB algorithm. We propose a two-phase policy called SE-RW-UCB: Subsidized Exploration with Retention-Weighted UCB. In the first T_{sub} rounds, the platform sets $r_t = 1$ and explores broadly so that viable providers accumulate a wealth buffer while the parameter estimates improve. It then switches to urgency-adjusted UCB ranking and online contract updates projected onto an estimated safe contract interval.

Algorithm 1: SE-RW-UCB

Input: Horizon T , subsidy horizon T_{sub} , feasible contract set $\mathcal{R}$, safety margin $\delta_{\min}$, step sizes $\{\eta_t\}_{t \geq 1}$

- 1 Initialize $r_0 \leftarrow 1$, $\mathcal{I}_1 = \mathcal{I}$, and $(V_{i,0}, \mathbf{b}_{i,0}, \bar{G}_{i,0}) \leftarrow (\lambda \mathbf{I}, \mathbf{0}, 0)$ for all i
- 2 **for** $t = 1, \dots, T$ **do**
- 3 **if** $t \leq T_{\text{sub}}$ **then**
- 4 **for** $\text{task } u \in \mathcal{U}_t$ **do**
- 5 Sample provider $i_t(u)$ uniformly from $\mathcal{I}_t$ and assign task u
- 6 $r_t \leftarrow 1$
- 7 **else**
- 8 **if** $t = T_{\text{sub}} + 1$ **then**
- 9 Estimate a safe contract floor $\hat{r}_{\text{safe}}$ from Phase-1 statistics over the surviving pool $\mathcal{I}_t$
- 10 Update $\hat{\theta}_{i,t-1}$, $\hat{c}_{i,t-1}$, $\hat{\mu}_{i,t}$, and $\Omega_{t,i}$ for all $i \in \mathcal{I}_t$
- 11 **for** $\text{task } u \in \mathcal{U}_t$ **do**
- 12 $\Omega_{t,i} \triangleq (1 - r_{t-1}) + N r_{t-1} \hat{\mu}_{i,t}$
- 13 $i_t(u) \leftarrow \arg \max_{i \in \mathcal{I}_t} \left\{ s_{ui} + \frac{\Omega_{t,i}}{\Lambda_i} \left(\mathbf{x}_u^\top \hat{\theta}_{i,t-1} + \beta_t \|\mathbf{x}_u\|_{V_{i,t-1}^{-1}} \right) \right\}$
- and assign task u
- 14 **for** $i \in \mathcal{I}_t$ **do**
- 15 $G_{i,\text{batch},t} \leftarrow \sum_{u: i_t(u)=i} y_{u,i,t}$
- 16 Update $(V_{i,t}, \mathbf{b}_{i,t})$, $\hat{\theta}_{i,t}$, $\hat{c}_{i,t}$, and $\bar{G}_{i,t}$
- 17 Observe $\mathbf{1}[i \in \mathcal{I}_{t+1}]$ and remove exited providers from the active set
- 18 **if** $t > T_{\text{sub}}$ **then**
- 19 $r_t \leftarrow \text{Proj}_{[\hat{r}_{\text{safe}}, 1]}(r_{t-1} + \eta_t \widehat{v}_r \mathcal{J}_t)$

Define the batch gradient estimator

$$\widehat{v}_r \mathcal{J}_t = -\frac{1}{N} G_{\text{batch},t} + \sum_{i \in \mathcal{I}_t} G_{i,\text{batch},t} \hat{\mu}_{i,t}, \quad (17)$$

where $G_{i,\text{batch},t}$ is provider i 's realized batch revenue, $G_{\text{batch},t} \triangleq \sum_{i \in \mathcal{I}_t} G_{i,\text{batch},t}$, and $\hat{\mu}_{i,t} = \Lambda_2 \sigma'(r_{t-1} \bar{G}_{i,t-1} - \hat{c}_{i,t-1})$ is the plug-in urgency weight computed from the running-average revenue $\bar{G}_{i,t-1}$ and the plug-in cost estimate $\hat{c}_{i,t-1} = c_{\text{base}} + \gamma \|\hat{\theta}_{i,t-1}\|_2$. Lemma A.2 shows that $|\hat{\mu}_{i,t} - \mu_i| = O(\|\hat{\theta}_{i,t-1} - \theta_i^*\|_2)$, so the urgency-estimation error inherits the standard least-squares rate. Under the linearized observation model with zero-mean noise, $\mathbb{E}[G_{i,\text{batch},t} \mid \mathcal{F}_{t-1}] = G_i(\pi_t)$ for every active provider i and every t . At the Phase-1/Phase-2 boundary, the algorithm computes a learned safe floor $\hat{r}_{\text{safe}}$ from Phase-1 statistics over the surviving pool. The Phase-2 analysis requires that this floor be accurate enough to remain below the oracle contract while still protecting viable providers on the confidence event introduced below. The multiplicative UCB score ensures that urgency cannot sustain a provider with negligible service capability, because the future-supply bonus always scales the reward estimate.

3.5 Theoretical Guarantees for SE-RW-UCB

We analyze the subsidy phase, the learned contract floor, and the post-subsidy ranking and ecosystem regret. Throughout, assume $\|\theta_i^*\|_2 \leq S$, $\|\mathbf{x}_u\|_2 \leq L$, and $0 \leq s_{ui} \leq 1$ for all i, u . Since $\sigma'(z) \leq \kappa/4$, we have $0 \leq \Omega_{t,i} \leq \Omega_{\max} \triangleq (1 - r_{\min}) + N \Lambda_2 \kappa/4$.

Safety of the subsidy phase. The subsidy phase decouples statistical discovery from financial survival.

Theorem 3.5 (Subsidy-phase survival and buffer accumulation). *Consider Phase 1 of Algorithm 1 with full subsidy $r_t = 1$ and uniform exploration. Suppose the Phase-1 net-revenue increments satisfy*

$$X_{i,t} \triangleq G_{i,\text{batch},t} - c_i, \quad |X_{i,t}| \leq B_{\text{rev}} \quad a.s.,$$

and

$$\mathbb{E}[X_{i,t} \mid \mathcal{F}_{t-1}] \geq \Delta_i > 0 \quad \text{for all } t \leq T_{\text{sub}}.$$

Fix a target post-subsidy wealth buffer $B_i \geq 0$ and define $W_{\text{safe}}^{(1)}(\delta) \triangleq \frac{1}{\kappa} \log \frac{16 T_{\text{sub}}}{\delta}$. If

$$W_{i,0} \geq W_{\text{safe}}^{(1)}(\delta) + \frac{B_{\text{rev}}^2}{\Delta_i} \log \frac{16}{\delta}$$

and

$$T_{\text{sub}} \geq \max \left\{ \frac{2(B_i - W_{i,0})_+}{\Delta_i}, \frac{32 B_{\text{rev}}^2}{\Delta_i^2} \log \frac{16}{\delta} \right\}, \quad (18)$$

then, with probability at least $1 - \delta$, provider i remains active throughout Phase 1 and satisfies $W_{i,T_{\text{sub}}} \geq B_i$.

Appendix A.6 contains the proof. Together with the lower bound (16), the theorem shows that safe exploration and buffer accumulation under the logistic departure law require both a nontrivial subsidy horizon and a positive initial wealth buffer. The post-subsidy statistical event $\mathcal{E}_\varepsilon(T)$ used in Theorem 3.6 is handled separately via a standard time-uniform self-normalized least-squares bound [1].

Ecosystem regret and conditional regret. To measure the post-subsidy phase, we separate within-pool assignment error from irreversible supply loss. Let $(r^*, \pi^*, \mathcal{I}^*)$ denote the oracle stationary policy under full knowledge of $\{\theta_i^*\}_{i \in \mathcal{I}}$ and the retained provider set. The ecosystem regret of Algorithm 1 over horizon T is

$$R_{\text{eco}}(T) = \sum_{t=1}^T \mathcal{W}(r^*, \pi^*, \mathcal{I}^*) - \sum_{t=1}^T \mathcal{W}(r_{t-1}, \pi_t, \mathcal{I}_t), \quad (19)$$

where $\mathcal{W}(r, \pi, \mathcal{I}') \triangleq \sum_u \sum_{i \in \mathcal{I}'} \rho_u \pi_{ui} [(1-r)R_{ui} + \Lambda_1 s_{ui}]$. Unlike ordinary bandit regret, $R_{\text{eco}}(T)$ penalizes both within-pool assignment mistakes and provider losses that the oracle would have avoided. For a tolerance $\varepsilon > 0$, define the confidence event $\mathcal{E}_\varepsilon(T) \triangleq \{\sup_{T_{\text{sub}} \leq t \leq T} \max_{i \in \mathcal{I}_t} \|\hat{\theta}_{i,t} - \theta_i^*\|_2 \leq \varepsilon\}$. Let $\mathcal{B}_\varepsilon(\theta^*) \triangleq \{\theta' \in \mathbb{R}^d : \|\theta' - \theta^*\|_2 \leq \varepsilon\}$ denote the ε -confidence ball around the true parameter. We define the *admissible policy set* $\Pi_c(\varepsilon)$ as the collection of all greedy assignment strategies compatible with parameters in this confidence region:

$$\Pi_c(\varepsilon) \triangleq \left\{ \pi \in \Pi : \exists \theta' \in \mathcal{B}_\varepsilon(\theta^*) \text{ s.t. } \pi = \arg \max_{\tilde{\pi} \in \Pi} \langle \mathbf{Z}(\tilde{\pi}; \theta'), \tilde{\pi} \rangle \right\}, \quad (20)$$

where $\mathbf{Z}(\cdot)$ denotes the assignment score vector defined in Line 13 of Algorithm 1. By construction, on the event $\mathcal{E}_\varepsilon(T)$ where the estimates are accurate, we have the guarantee: $\{\pi_t : t > T_{\text{sub}}\} \subseteq \Pi_c(\varepsilon)$. This allows us to bound the worst-case provider revenue under accurate learning: $G_i^{\min}(\varepsilon) \triangleq \inf_{\pi \in \Pi_c(\varepsilon)} G_i(\pi)$. Define the oracle safe contract floor $r_{\text{safe}} \triangleq \max_{i \in \mathcal{I}^*} \frac{c_i + 2\delta_{\min}}{G_i^{\min}(\varepsilon)}$.

Let $\hat{r}_{\text{safe}}$ denote the learned floor used in Algorithm 1. On $\mathcal{E}_\varepsilon(T)$, the post-subsidy policies induced by Algorithm 1 satisfy $\pi_t \in \Pi_c(\varepsilon)$ for all $t > T_{\text{sub}}$. We assume: **(i)** $r^* G_i(\pi^*) - c_i \geq 3\delta_{\min}$ for all $i \in \mathcal{I}^*$; **(ii)** $G_i^{\min}(\varepsilon) > 0$ for all $i \in \mathcal{I}^*$; **(iii)** $r_{\text{safe}} < r^*$; **(iv)** $r^* G_i(\pi^*) - c_i \geq \Delta_{\text{gap}}$ for $i \in \mathcal{I}^*$, while $r^* G_j(\pi^*) - c_j \leq -\Delta_{\text{gap}}$ for $j \notin \mathcal{I}^*$; and **(v)** on

$\mathcal{E}_\epsilon(T)$, the learned floor satisfies $|\hat{r}_{\text{safe}} - r_{\text{safe}}| \leq C_{\text{safe}}\epsilon$. This setup only requires viability above one threshold, which the projection step enforces through the learned safe interval.

In the regret analysis below, we choose ϵ small enough that

$$C_{\text{safe}}\epsilon\tilde{G}_{\text{max}} \leq \frac{\delta_{\min}}{2} \quad \text{and} \quad r_{\text{safe}} + C_{\text{safe}}\epsilon < r^*.$$

Theorem 3.6 (Ecosystem regret bound). *Let V_{max} be a finite upper bound on the one-period welfare contribution of a single provider and define $G_{\text{max}} \triangleq \max_{\pi} \sum_{i \in \mathcal{I}} G_i(\pi)$. Assume there exists $G_{\text{max},\text{batch}}$ such that $|G_{i,\text{batch},t}| \leq G_{\text{max},\text{batch}}$ almost surely for every i, t , and define $B_{\text{rev}} \triangleq G_{\text{max},\text{batch}} + \max_i c_i$. Let*

$$B^* \triangleq \frac{1}{\kappa} \log \frac{4(T - T_{\text{sub}})|\mathcal{I}^*|}{\delta} + \frac{2B_{\text{rev}}^2}{\delta_{\min}} \log \frac{4|\mathcal{I}^*|}{\delta}.$$

Suppose the post-subsidy setup above holds, ϵ is chosen as above, the uniform post-subsidy confidence event satisfies $\mathbb{P}(\mathcal{E}_\epsilon(T)) \geq 1 - \delta/4$, and Theorem 3.5 applies to every $i \in \mathcal{I}^*$ with target buffer $B_i = B^*$ and failure probability $\delta/(4|\mathcal{I}^*|)$. Then, with probability at least $1 - \delta$, every oracle-viable provider survives Phase 2, that is, $\mathcal{I}^* \subseteq \mathcal{I}_t$ for all $t > T_{\text{sub}}$, and the ecosystem regret satisfies

$$R_{\text{eco}}(T) \leq T_{\text{sub}}|\mathcal{I}|V_{\text{max}} + R(T) + \frac{G_{\text{max}}}{N} \sum_{t=T_{\text{sub}}+1}^T |r_{t-1} - r^*|, \quad (21)$$

where $R(T)$ is the conditional regret defined in (22).

Appendix A.7 proves the ecosystem bound.

Conditional regret. Define the post-subsidy conditional regret by

$$R(T) = \sum_{t=T_{\text{sub}}+1}^T (Z_{t,i_t^*}^* - Z_{t,i_t}^*). \quad (22)$$

Here $Z_{t,k}^* \triangleq s_{t,k} + (\Omega_{t,k}/\Lambda_1)\langle \theta_k^*, \mathbf{x}_{t,k} \rangle$ and $i_t^* \in \arg \max_{k \in \mathcal{I}_t} Z_{t,k}^*$. Standard OFUL analysis [1] in Appendix A.8 gives

$$R(T) \leq \frac{2\Omega_{\text{max}}}{\Lambda_1} \beta_T \sqrt{2|\mathcal{I}|d(T - T_{\text{sub}}) \log \left(1 + \frac{(T - T_{\text{sub}})L^2}{\lambda d} \right)},$$

with probability at least $1 - \delta$, where β_T is the standard OFUL confidence radius. This is the usual $\tilde{O}(\sqrt{|\mathcal{I}|dT})$ linear-bandit rate for $|\mathcal{I}|$ separate provider models, scaled by the uniform urgency bound.

4 Experiments

This section evaluates the SE-RW-UCB algorithm empirically. We ask three questions. First, does joint contract–recommendation design outperform each lever alone? Second, is the low-contract synthetic regime a true interior optimum or a floor-driven boundary effect? Third, which ingredients of SE-RW-UCB are empirically responsible for its gains?

4.1 Experimental Protocol

The semi-synthetic KuaiRec benchmark [8] uses 200 sampled users, 50 retained creators, and 50 user groups at horizon $T = 5,000$. In the comparison, every contract-optimizing method operates on the common feasible interval $\mathcal{R} = [0.1, 1]$, matching the constrained theory in Section 3.2. The main semi-synthetic table averages over 10 seeds. All intervals are 95% bootstrap confidence intervals computed from seed-level summaries.

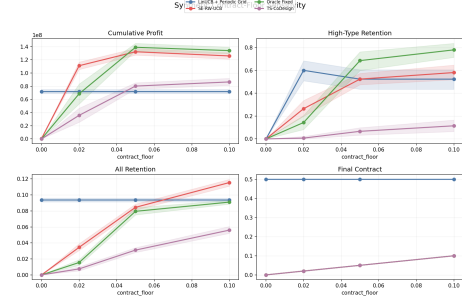


Figure 1: Synthetic contract-floor sensitivity. Oracle Fixed and SE-RW-UCB track the imposed floor when it is relaxed from 0.10 to 0.05 and 0.02, revealing that the synthetic low-contract regime is boundary-driven rather than interior. Raising the floor from 0.05 to 0.10 sacrifices some profit but improves both high-type and all-creator retention.

Baselines. We compare SE-RW-UCB against six alternatives.

- (1) **Greedy** ($r = 0.5$): fixed contract with myopic revenue ranking.
- (2) **Static Subsidy** ($r = 0.8$): permanently generous fixed contract with myopic ranking.
- (3) **Oracle Fixed**: hindsight-optimal fixed contract over the constrained grid.
- (4) **LinUCB + Periodic Grid**: contextual ranking [19] with contract re-optimization every 500 rounds.
- (5) **TS-CoDesign**: Thompson-sampling co-design rule.
- (6) **Conservative Safe Bandit (CSB)**: fixed-contract contextual ranking with minimum-exposure safeguards.

Metrics. The main comparisons report cumulative platform profit, cumulative user satisfaction, high-type retention, all-creator retention, and the final contract. Profit and user satisfaction are reported in millions.

4.2 Semi-Synthetic Validation on KuaiRec

Table 1 answers the first question. Oracle Fixed attains the highest mean cumulative profit, but SE-RW-UCB leads user satisfaction, high-type retention, and all-creator retention simultaneously. LinUCB + Periodic Grid remains competitive on creator preservation, yet it is uniformly weaker than SE-RW-UCB on the user-facing and creator-facing metrics. The 10 sampled seeds contain between 10 and 15 high-type creators, with mean 12.3, so the retention intervals remain wide.

4.3 Ablation Study

Figure 1 answers the second question. The low-contract regime is boundary-driven: when the feasible set is relaxed from $[0.1, 1]$ to $[0.05, 1]$ and to $[0.02, 1]$, both Oracle Fixed and SE-RW-UCB move to the new lower boundary. At $r_{\min} = 0$, the adaptive co-design rules collapse to zero contracts and near-zero retention, whereas at $r_{\min} = 0.1$ they retain substantially broader creator pools. The floor therefore parameterizes a profit–retention trade-off in the synthetic calibration: lowering the floor raises profit, but sharply reduces both high-type and all-creator retention.

Table 1: Semi-synthetic KuaiRec comparison over $T = 5,000$ rounds with 10 random seeds. Profit and user satisfaction are reported in millions; entries are means with 95% bootstrap confidence intervals.

Algorithm	Cumul. Profit	Cumul. User Sat.	High-Type Ret.	All Ret.	Final Contract
Greedy	0.38 [0.36, 0.41]	0.713 [0.705, 0.722]	0.059 [0.033, 0.083]	0.020 [0.020, 0.020]	0.500 [0.500, 0.500]
Static Subsidy	0.15 [0.14, 0.16]	0.713 [0.705, 0.722]	0.059 [0.033, 0.083]	0.020 [0.020, 0.020]	0.800 [0.800, 0.800]
Oracle Fixed	1.66 [1.59, 1.74]	0.781 [0.770, 0.790]	0.041 [0.015, 0.066]	0.140 [0.114, 0.162]	0.170 [0.140, 0.200]
LinUCB + Periodic Grid	1.10 [1.07, 1.14]	0.783 [0.776, 0.789]	0.216 [0.164, 0.265]	0.290 [0.272, 0.308]	0.500 [0.500, 0.500]
TS-CoDesign	1.32 [1.22, 1.42]	0.793 [0.780, 0.806]	0.092 [0.046, 0.141]	0.088 [0.078, 0.098]	0.102 [0.100, 0.104]
Conservative Safe Bandit	0.72 [0.69, 0.74]	0.737 [0.731, 0.744]	0.247 [0.155, 0.335]	0.278 [0.258, 0.294]	0.500 [0.500, 0.500]
SE-RW-UCB	1.48 [1.29, 1.68]	0.822 [0.815, 0.828]	0.279 [0.187, 0.379]	0.338 [0.258, 0.412]	0.264 [0.182, 0.347]

Table 2 answers the third question by isolating which ingredients matter most on the semi-synthetic benchmark. Removing the subsidy phase or fixing the contract after warm start lowers all retention from roughly 0.34 to the 0.16–0.19 range. Warm-starting TS-CoDesign recovers nearly the same broad retention as SE-RW-UCB, showing that the temporary subsidy phase is a major part of the creator-preservation effect. SE-RW-UCB achieves the highest mean high-type retention and remains on the top user-satisfaction frontier, so the RW-UCB structure sharpens which creators are preserved rather than merely increasing the total retained pool.

5 Related Work

Classical platform theory [2, 4, 6, 7, 26, 27] studies price structure across market sides, while recent work on algorithm-curated platforms [11, 12, 16, 35, 36] shows that recommendation and monetization policies reshape the provider pool. Our model adds an explicit contract lever and a mobile-systems interpretation covering crowdsensing, edge computing, vehicular services, and sponsored content within one framework.

In mobile services, incentive design has largely treated assignment and payment separately. Truthful crowdsensing pricing [10, 34], competition among crowdsourcers [5], auction-based edge offloading and resource sharing [24, 29, 38, 40], contextual bandits [1, 19], safe-bandit variants [9, 14, 25, 32], and wireless bandit models [20, 39] all provide ingredients for our setting, but they typically assume a fixed resource pool. Relative to dynamic-population models [21] and recommendation learning with user self-selection [28], our framework couples contextual heterogeneity, explicit contracting, irreversible departure, and ecosystem-level regret in a mobile-platform environment.

6 Conclusion

We study mobile networked platforms with endogenous provider supply and show that assignment and incentive contracts are substitute retention levers that should be optimized in common welfare units. This yields a surrogate-based calibration of the stationary objective, a complete-information benchmark with finite-time joint stationarity, and a learning algorithm with safe exploration and a learned contract floor. In the KuaiRec benchmark, Oracle Fixed leads profit, while SE-RW-UCB delivers the strongest high-type and all-provider retention with an interior contract. In this calibration, the main benefit of co-design is supply preservation rather than per-request quality improvement.

A Deferred Proofs and Auxiliary Results

A.1 Global Conservative Lower-Bound Calibration

Proposition A.1 (Global conservative lower bound for the auxiliary surrogate). *Fix a compact feasible set $\mathcal{K}$ on which $w_i(r, \pi) \geq 0$ for every provider i . Let $\bar{w}^\circ > 0$ be any positive reference welfare level and define $B_w \triangleq \sup_{(r, \pi) \in \mathcal{K}, i \in I} |w_i(r, \pi) - \bar{w}^\circ|$. Then, for every $(r, \pi) \in \mathcal{K}$,*

$$\mathcal{J}_{\text{aux}}(r, \pi) \geq \mathcal{W}(r, \pi) + \delta \bar{w}^\circ S(r, \pi) - \delta |I| B_w.$$

Since the additive constant $\delta |I| B_w$ does not depend on (r, π) , any maximizer of the surrogate (8) with $\Lambda_2 = \delta \bar{w}^\circ$ simultaneously maximizes this global lower bound on $\mathcal{J}_{\text{aux}}$.

PROOF. Let $p_i(r, \pi) = \sigma(rG_i(\pi) - c_i)$. Since $(1 - x)^{-1} \geq 1 + x$ on $x \in [0, 1)$, we have

$$\begin{aligned} \mathcal{J}_{\text{aux}}(r, \pi) &= \sum_i \frac{w_i(r, \pi)}{1 - \delta p_i(r, \pi)} \\ &\geq \sum_i w_i(r, \pi) (1 + \delta p_i(r, \pi)) \\ &= \mathcal{W}(r, \pi) + \delta \sum_i w_i(r, \pi) p_i(r, \pi). \end{aligned}$$

Write $w_i(r, \pi) = \bar{w}^\circ + (w_i(r, \pi) - \bar{w}^\circ)$. Then

$$\sum_i w_i(r, \pi) p_i(r, \pi) = \bar{w}^\circ S(r, \pi) + \sum_i (w_i(r, \pi) - \bar{w}^\circ) p_i(r, \pi).$$

Because $0 \leq p_i(r, \pi) \leq 1$ and $|w_i(r, \pi) - \bar{w}^\circ| \leq B_w$ on $\mathcal{K}$,

$$\sum_i (w_i(r, \pi) - \bar{w}^\circ) p_i(r, \pi) \geq -|I| B_w.$$

Substituting establishes the claim. $\square$

A.2 Proof of Proposition 2.1

PROOF. Differentiating (9) yields

$$\nabla \mathcal{J}_{\text{aux}} = \sum_i \left[\frac{\nabla w_i}{1 - \delta p_i} + \frac{\delta w_i}{(1 - \delta p_i)^2} \nabla p_i \right].$$

By compactness of $\mathcal{K}$, gradients $\nabla w_i, \nabla p_i$ and weights w_i are uniformly bounded. Since $p_i^\circ \geq 1 - \varepsilon$, first-order expansions give

$$\left| \frac{1}{1 - \delta p_i^\circ} - \frac{1}{1 - \delta} \right| \leq \frac{\delta \varepsilon}{(1 - \delta)^2},$$

$$\left| \frac{1}{(1 - \delta p_i^\circ)^2} - \frac{1}{(1 - \delta)^2} \right| \leq \frac{2\delta \varepsilon}{(1 - \delta)^3}.$$

Using $|w_i^\circ - \bar{w}^\circ| \leq \eta$ and $\nabla W = \sum_i \nabla w_i, \nabla S = \sum_i \nabla p_i$, the bound follows with constants absorbing $(1 - \delta)^{-3}$ and derivative maxima. $\square$

Table 2: Semi-synthetic ablations over $T = 5,000$ rounds with 10 random seeds. Profit and user satisfaction are reported in millions; entries are means with 95% bootstrap confidence intervals.

Algorithm	Cumul. Profit	Cumul. User Sat.	High-Type Ret.	All Ret.	Final Contract
Subsidy-Only-UCB	1.73 [1.63, 1.86]	0.801 [0.784, 0.814]	0.135 [0.062, 0.217]	0.192 [0.164, 0.222]	0.130 [0.111, 0.151]
Subsidy+Urgency-NoContractLearn	1.66 [1.56, 1.80]	0.822 [0.813, 0.830]	0.174 [0.099, 0.260]	0.258 [0.220, 0.296]	0.170 [0.140, 0.200]
FixedContract-RW-UCB	1.55 [1.45, 1.66]	0.799 [0.785, 0.815]	0.131 [0.092, 0.163]	0.162 [0.126, 0.196]	0.170 [0.140, 0.200]
NoSubsidy-RW-UCB	1.48 [1.39, 1.58]	0.815 [0.803, 0.829]	0.175 [0.131, 0.220]	0.158 [0.132, 0.190]	0.132 [0.105, 0.168]
SE-RW-UCB	1.48 [1.29, 1.68]	0.822 [0.815, 0.828]	0.279 [0.187, 0.379]	0.338 [0.258, 0.412]	0.264 [0.182, 0.347]
WarmStart-TS-CoDesign	1.42 [1.24, 1.63]	0.816 [0.809, 0.822]	0.271 [0.184, 0.372]	0.340 [0.270, 0.404]	0.269 [0.184, 0.368]
LinUCB + Periodic Grid	1.10 [1.07, 1.14]	0.783 [0.776, 0.789]	0.216 [0.164, 0.265]	0.290 [0.272, 0.308]	0.500 [0.500, 0.500]

A.3 Proof of Proposition 3.3

PROOF. Fix $r \in \mathcal{R}$. The partial derivative of $\mathcal{J}$ with respect to π_{ui} is $\frac{\partial \mathcal{J}}{\partial \pi_{ui}} = \rho_u [\Lambda_1 s_{ui} + R_{ui}((1-r) + Nr\mu_i(\pi))]$, where $\mu_i(\pi) = \Lambda_2 \sigma'(rG_i(\pi) - c_i)$. At a stationary assignment π^* , this is exactly $\rho_u \Lambda_1 Z_{ui}^*$ with Z_{ui}^* from (13). The simplex KKT conditions therefore imply $\pi_{ui}^* > 0 \Rightarrow Z_{ui}^* \geq Z_{uj}^*$ for all j , which proves Proposition 3.3.

Scalar contract KKT condition. For the contract,

$$\begin{aligned} \partial_r \mathcal{J}(r, \pi) &= -\frac{1}{N} \sum_i G_i(\pi) \\ &\quad + \Lambda_2 \sum_i G_i(\pi) \sigma'(rG_i(\pi) - c_i) \triangleq g_r(r, \pi). \end{aligned}$$

The projected KKT condition at (r^*, π^*) is $g_r(r^*, \pi^*)(r - r^*) \leq 0$ for all $r \in \mathcal{R}$, namely

$$g_r(r^*, \pi^*)(r - r^*) \leq 0 \quad \forall r \in \mathcal{R}. \quad (23)$$

When r^* is interior, this reduces to $g_r(r^*, \pi^*) = 0$, namely

$$\frac{1}{N} \sum_i G_i(\pi^*) = \Lambda_2 \sum_i G_i(\pi^*) \sigma'(r^* G_i(\pi^*) - c_i). \quad (24)$$

□

A.4 Proof of Proposition 3.4

PROOF. First fix $r \in \mathcal{R}$. Since $\mathcal{W}(r, \pi)$ is affine in π , only the retention term contributes to the Hessian. For any direction $d \in \mathbb{R}^{|\mathcal{U}| \times |\mathcal{I}|}$,

$$d^\top \nabla_\pi^2 [\Lambda_2 S(r, \pi)] d = \Lambda_2 \sum_i \sigma''(rG_i(\pi) - c_i) \left(\sum_u r N \rho_u R_{ui} d_{ui} \right)^2.$$

Using $|\sigma''(z)| \leq \kappa^2/4$, Cauchy–Schwarz, and $\sum_u \rho_u^2 \leq 1$ gives

$$|d^\top \nabla_\pi^2 [\Lambda_2 S(r, \pi)] d| \leq \frac{\Lambda_2 \kappa^2}{4} N^2 R_{\max}^2 \|d\|^2 = L_\pi \|d\|^2.$$

Hence $\|\nabla_\pi^2 \mathcal{J}(r, \pi)\|_{\text{op}} \leq L_\pi$, so $\pi \mapsto \mathcal{J}(r, \pi)$ has L_π -Lipschitz gradient on Π . Because Π is a product of $|\mathcal{U}|$ simplices, its squared Euclidean diameter is $D_\Pi^2 = 2|\mathcal{U}|$, and therefore $C_\Pi = L_\pi D_\Pi^2$.

Set

$$\mathcal{J}^{(t)} \triangleq \mathcal{J}(r^{(t)}, \pi^{(t)}), \quad g_t \triangleq \mathcal{G}_\eta(r^{(t)}, \pi^{(t+1)}).$$

Because the feasible set $\mathcal{R} \times \Pi$ is compact and $\mathcal{J}$ is continuous, $\mathcal{J}^* = \sup_{r, \pi} \mathcal{J}(r, \pi)$ is finite.

Assignment and contract steps. Standard nonconvex Frank–Wolfe ascent [17] with $\gamma_t = \min\{1, \text{FWgap}_t/C_\Pi\}$ yields

$$\mathcal{J}(r^{(t)}, \pi^{(t+1)}) - \mathcal{J}(r^{(t)}, \pi^{(t)}) \geq \min\left\{\frac{\text{FWgap}_t^2}{2C_\Pi}, \frac{C_\Pi}{2}\right\}.$$

Since $r \mapsto \mathcal{J}(r, \pi^{(t+1)})$ is L_r -smooth and $0 < \eta \leq 1/L_r$, standard projected-gradient ascent gives

$$\mathcal{J}(r^{(t+1)}, \pi^{(t+1)}) - \mathcal{J}(r^{(t)}, \pi^{(t+1)}) \geq \frac{\eta}{2} \|g_t\|^2.$$

Telescope. Summing $\mathcal{J}^{(t+1)} - \mathcal{J}^{(t)} \geq \frac{\eta}{2} \|g_t\|^2$ from $t = 0$ to $K - 1$ yields

$$\min_{0 \leq t \leq K-1} \|g_t\|^2 \leq \frac{2(\mathcal{J}^* - \mathcal{J}^{(0)})}{\eta K}.$$

For the Frank–Wolfe gap, split $\{0, \dots, K-1\}$ into $\mathcal{A} = \{t : \text{FWgap}_t > C_\Pi\}$ and $\mathcal{B}$ its complement. By the assignment step,

$$|\mathcal{A}| \leq \frac{2(\mathcal{J}^* - \mathcal{J}^{(0)})}{C_\Pi}, \quad \sum_{t \in \mathcal{B}} \text{FWgap}_t^2 \leq 2C_\Pi(\mathcal{J}^* - \mathcal{J}^{(0)}).$$

If $K > 4(\mathcal{J}^* - \mathcal{J}^{(0)})/C_\Pi$, then $|\mathcal{B}| > K/2$, and therefore

$$\min_{0 \leq t \leq K-1} \text{FWgap}_t^2 \leq \frac{4C_\Pi(\mathcal{J}^* - \mathcal{J}^{(0)})}{K}.$$

This proves the claimed $O(1/K)$ joint stationarity rate. □

A.5 Information-Floor Lower Bound

PROOF OF (16). The platform must distinguish the two hypotheses $H_0 : \theta = \theta_H$ and $H_1 : \theta = \theta_L$ from noisy observations

$$y_t = \theta + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \sigma^2).$$

Let $\mathbb{P}_H^{(T)}$ and $\mathbb{P}_L^{(T)}$ denote the laws of the first T observations under the two hypotheses. Because $\theta_H - \theta_L = 2\Delta$,

$$D_{\text{KL}}(\mathbb{P}_H^{(T)} \parallel \mathbb{P}_L^{(T)}) = \frac{T(\theta_H - \theta_L)^2}{2\sigma^2} = \frac{2T\Delta^2}{\sigma^2}.$$

Bretagnolle–Huber’s inequality (see, e.g., [18]) yields

$$\mathbb{P}_H(\hat{\theta} \neq \theta_H) + \mathbb{P}_L(\hat{\theta} \neq \theta_L) \geq \frac{1}{2} \exp(-D_{\text{KL}}(\mathbb{P}_H^{(T)} \parallel \mathbb{P}_L^{(T)})).$$

Hence

$$\max\{\mathbb{P}_H(\hat{\theta} \neq \theta_H), \mathbb{P}_L(\hat{\theta} \neq \theta_L)\} \geq \frac{1}{4} \exp\left(-\frac{2T\Delta^2}{\sigma^2}\right).$$

If the worst-case classification error is at most $\delta < 1/4$, then $\delta \geq \frac{1}{4} \exp\left(-\frac{2T\Delta^2}{\sigma^2}\right)$, which rearranges to (16). The same lower bound applies to the full linear contextual model by restricting to the one-dimensional subclass $d = 1$, $\mathbf{x}_u \equiv 1$, and $\theta_i^* \in \{\theta_L, \theta_H\}$. □

Lemma A.2 (Shadow-price estimation error). Fix $r \in \mathcal{R}$ and π . Define the oracle urgency $\mu_i = \Lambda_2 \sigma'(rG_i - c_i)$ and the plug-in estimate

$$\hat{\mu}_i = \Lambda_2 \sigma'(r\hat{G}_i - \hat{c}_i), \quad \hat{G}_i = N \sum_u \rho_u \pi_{ui} \langle \hat{\theta}_i, \mathbf{x}_u \rangle.$$

If $\varepsilon_i \triangleq \|\hat{\theta}_i - \theta_i^*\|_2$, then

$$|\hat{\mu}_i - \mu_i| \leq \Lambda_2 \frac{\kappa^2}{4} (rNL + \gamma) \varepsilon_i. \quad (25)$$

PROOF. Because $\sigma(z) = (1 + e^{-\kappa z})^{-1}$,

$$\sigma''(z) = \kappa \sigma'(z) (1 - 2\sigma(z)).$$

Since $\sigma'(z) \leq \kappa/4$ for all z , it follows that $|\sigma''(z)| \leq \kappa^2/4$. Hence σ' is Lipschitz with constant $\kappa^2/4$.

Write $a_i = rG_i - c_i$ and $\hat{a}_i = r\hat{G}_i - \hat{c}_i$. Then

$$|\hat{\mu}_i - \mu_i| = \Lambda_2 |\sigma'(\hat{a}_i) - \sigma'(a_i)| \leq \Lambda_2 \frac{\kappa^2}{4} |\hat{a}_i - a_i|.$$

It remains to bound $|\hat{a}_i - a_i|$. First,

$$|\hat{G}_i - G_i| = N \left| \sum_u \rho_u \pi_{ui} \langle \hat{\theta}_i - \theta_i^*, \mathbf{x}_u \rangle \right| \leq NL\epsilon_i \sum_u \rho_u \pi_{ui} \leq NL\epsilon_i,$$

using $\sum_u \rho_u \pi_{ui} \leq \sum_u \rho_u = 1$. Second,

$$|\hat{c}_i - c_i| = \gamma \left| \|\hat{\theta}_i\|_2 - \|\theta_i^*\|_2 \right| \leq \gamma \epsilon_i.$$

Therefore,

$$|\hat{a}_i - a_i| \leq r|\hat{G}_i - G_i| + |\hat{c}_i - c_i| \leq (rNL + \gamma)\epsilon_i,$$

which proves (25). $\square$

A.6 Proof of Theorem 3.5

PROOF. Let

$$S_n \triangleq \sum_{\tau=1}^n X_{i,\tau}, \quad W_{i,n} = W_{i,0} + S_n.$$

Part 1: pathwise survival above a safety threshold. Set

$$a \triangleq W_{i,0} - W_{\text{safe}}^{(1)}(\delta), \quad \mu \triangleq \Delta_i, \quad b \triangleq B_{\text{rev}}.$$

Because $a \geq (B_{\text{rev}}^2/\Delta_i) \log(16/\delta)$ by assumption, Lemma A.3 gives

$$\begin{aligned} \mathbb{P}\left(\inf_{1 \leq n \leq T_{\text{sub}}} W_{i,n} \leq W_{\text{safe}}^{(1)}(\delta)\right) &\leq \mathbb{P}\left(\inf_{n \geq 1} S_n \leq -a\right) \\ &\leq \exp\left(-\frac{a\Delta_i}{B_{\text{rev}}^2}\right) \leq \frac{\delta}{16}. \end{aligned}$$

Part 2: no exit while above the threshold. On the complementary event, $W_{i,t} \geq W_{\text{safe}}^{(1)}(\delta)$ for every $t \leq T_{\text{sub}}$. Hence

$$1 - \sigma(W_{i,t}) \leq 1 - \sigma(W_{\text{safe}}^{(1)}(\delta)) \leq e^{-\kappa W_{\text{safe}}^{(1)}(\delta)} = \frac{\delta}{16T_{\text{sub}}}.$$

A union bound over the T_{sub} subsidy rounds shows that the probability of any departure during Phase 1 while the wealth path stays above the threshold is at most $\delta/16$. Therefore provider i survives all of Phase 1 with probability at least $1 - \delta/8$.

Part 3: terminal buffer accumulation. The centered martingale differences $Y_{i,t} = X_{i,t} - \mathbb{E}[X_{i,t} | \mathcal{F}_{t-1}]$ satisfy $|Y_{i,t}| \leq 2B_{\text{rev}}$, so Azuma–Hoeffding gives

$$\mathbb{P}\left(W_{i,T_{\text{sub}}} < W_{i,0} + \frac{T_{\text{sub}}\Delta_i}{2}\right) \leq \exp\left(-\frac{T_{\text{sub}}\Delta_i^2}{32B_{\text{rev}}^2}\right) \leq \frac{\delta}{16}.$$

Since $T_{\text{sub}} \geq 2(B_i - W_{i,0})/\Delta_i$, the complement of the event above implies $W_{i,T_{\text{sub}}} \geq B_i$, so

$$\mathbb{P}(W_{i,T_{\text{sub}}} \geq B_i) \geq 1 - \frac{\delta}{16}.$$

A union bound over Parts 1–3 yields the claim. $\square$

Lemma A.3 (Ruin probability; standard). *Let $(X_n)_{n \geq 1}$ be adapted and define $S_n = \sum_{k=1}^n X_k$. If $|X_n| \leq b$ almost surely and $\mathbb{E}[X_n | \mathcal{F}_{n-1}] \geq \mu > 0$ for all n , then, for every $a > 0$,*

$$\mathbb{P}\left(\inf_{n \geq 1} S_n \leq -a\right) \leq \exp\left(-\frac{a\mu}{b^2}\right).$$

PROOF. Set $\lambda = \mu/b^2$. Hoeffding's lemma gives

$$\mathbb{E}\left[e^{-\lambda X_n} | \mathcal{F}_{n-1}\right] \leq \exp\left(-\frac{\mu^2}{2b^2}\right),$$

so $M_n \triangleq \exp\left(-\lambda S_n + \frac{n\mu^2}{2b^2}\right)$ is a non-negative supermartingale. Applying optional stopping at $\tau = \inf\{n \geq 1 : S_n \leq -a\}$ yields

$$\mathbb{P}(\tau < \infty) \leq \exp\left(-\frac{a\mu}{b^2}\right).$$

This is the standard exponential-supermartingale argument. $\square$

A.7 Proof of Theorem 3.6

PROOF. The argument has three parts.

Part 1: Phase-1 buffer and post-subsidy confidence. Let

$$B^* \triangleq \frac{1}{\kappa} \log \frac{4(T - T_{\text{sub}})|\mathcal{I}^*|}{\delta} + \frac{2B_{\text{rev}}^2}{\delta_{\min}} \log \frac{4|\mathcal{I}^*|}{\delta}$$

and apply Theorem 3.5 to each $i \in \mathcal{I}^*$ with target buffer $B_i = B^*$ and failure probability $\delta/(4|\mathcal{I}^*|)$. By a union bound, with probability at least $1 - \delta/4$, every oracle-viable provider survives Phase 1 and enters Phase 2 with wealth at least B^* . On the event $\mathcal{E}_\epsilon(T)$, which holds with probability at least $1 - \delta/4$ by assumption, every post-subsidy policy lies in Π_ϵ and the learned floor satisfies $|\hat{r}_{\text{safe}} - r_{\text{safe}}| \leq C_{\text{safe}}\epsilon$. Thus, with probability at least $1 - \delta/2$, every oracle-viable provider enters Phase 2 alive with wealth at least B^* , every post-subsidy policy lies in Π_ϵ , and $\hat{r}_{\text{safe}} < r^*$.

Part 2: Phase-2 survival. Fix $i \in \mathcal{I}^*$ and condition on the joint event from Part 1. Define the post-subsidy filtration $\mathcal{H}_n \triangleq \mathcal{F}_{T_{\text{sub}}+n}$, and the Phase-2 increments

$$X_n \triangleq r_{T_{\text{sub}}+n-1}G_{i,\text{batch},T_{\text{sub}}+n} - c_i, \quad n \geq 1.$$

Construct auxiliary increments $\tilde{X}_n$ that coincide with X_n while provider i is active and equal $\delta_{\min}/2$ thereafter. Then $|\tilde{X}_n| \leq B_{\text{rev}}$ a.s. For each round $t > T_{\text{sub}}$, the batch is served under contract r_{t-1} , the projection step ensures $r_{t-1} \geq \hat{r}_{\text{safe}}$, and $\pi_t \in \Pi_\epsilon(\epsilon)$ implies $G_i(\pi_t) \geq G_i^{\min}(\epsilon)$. By assumption (v), $\hat{r}_{\text{safe}} \geq r_{\text{safe}} - C_{\text{safe}}\epsilon$. Hence

$$\begin{aligned} r_{t-1}G_i(\pi_t) - c_i &\geq (r_{\text{safe}} - C_{\text{safe}}\epsilon)G_i^{\min}(\epsilon) - c_i \\ &\geq 2\delta_{\min} - C_{\text{safe}}\epsilon \max_{i,\pi} G_i(\pi) \geq \frac{3}{2}\delta_{\min}. \end{aligned}$$

By the conditional-unbiasedness assumption stated after (17), $\mathbb{E}[G_{i,\text{batch},t} | \mathcal{F}_{t-1}] = G_i(\pi_t)$ for every active provider i . Therefore $\mathbb{E}[\tilde{X}_n | \mathcal{H}_{n-1}] \geq \frac{\delta_{\min}}{2}$ for each Phase-2 increment. Applying Lemma A.3 with $a = B^* - W_0$, $\mu = \delta_{\min}/2$, and $b = B_{\text{rev}}$, where $W_0 \triangleq \frac{1}{\kappa} \log \frac{4(T - T_{\text{sub}})|\mathcal{I}^*|}{\delta}$, gives the per-provider ruin bound

$$\mathbb{P}\left(\inf_{n \geq 1} \sum_{k=1}^n \tilde{X}_k \leq -(B^* - W_0)\right) \leq \exp\left(-\frac{(B^* - W_0)\delta_{\min}}{2B_{\text{rev}}^2}\right) = \frac{\delta}{4|\mathcal{I}^*|}.$$

On the complementary event, the true wealth of provider i always exceeds W_0 while it remains active. Therefore the period-by-period departure probability is at most $1 - \sigma(W_0) \leq e^{-\kappa W_0} = \frac{\delta}{4(T - T_{\text{sub}})|\mathcal{I}^*|}$. A union bound over rounds shows that provider i exits in Phase 2 with probability at most $\delta/(2|\mathcal{I}^*|)$. Union-bounding again over all $i \in \mathcal{I}^*$ yields

$$\mathbb{P}(\exists i \in \mathcal{I}^*, \exists t > T_{\text{sub}} \text{ with } i \notin \mathcal{I}_t) \leq \frac{\delta}{2}.$$

Part 3: Ecosystem regret decomposition. Adding and subtracting $\sum_{t=T_{\text{sub}}+1}^T \mathcal{W}(r^*, \pi^*, I_t)$ inside (19) and splitting the first T_{sub} rounds gives $R_{\text{eco}}(T) = R_{\text{sub}}(T) + R_{\text{att}}(T) + R_{\text{pol}}(T)$. On the joint event from Parts 1 and 2, all providers in $\mathcal{I}^*$ survive Phase 2. Hence $R_{\text{sub}}(T) \leq T_{\text{sub}}|J|V_{\max}$, $R_{\text{att}}(T) \leq 0$, and

$$R_{\text{pol}}(T) \leq R(T) + \frac{G_{\max}}{N} \sum_{t=T_{\text{sub}}+1}^T |r_{t-1} - r^*|.$$

Combining the three bounds proves (21). The total failure probability is at most δ . $\square$

A.8 Proof of the Conditional Regret Bound

PROOF. Condition on the standard OFUL event holding with probability $\geq 1 - \delta$. Let $\Gamma_{t,i} = \Omega_{t,i}/\Lambda_1$. Optimism implies $Z_{t,i_t}^{\text{UCB}} \geq Z_{t,i_t^*}^{\text{UCB}} \geq Z_{t,i_t^*}^*$, giving instantaneous regret $\leq 2\Gamma_{t,i_t}\beta_t\|\mathbf{x}_{t,i_t}\|_{V_{t,t-1}^{-1}}$. Summing and using $\Gamma_{t,i} \leq \Omega_{\max}/\Lambda_1$ yields

$$R(T) \leq \frac{2\Omega_{\max}}{\Lambda_1}\beta_T \sum_{t=T_{\text{sub}}+1}^T \|\mathbf{x}_{t,i_t}\|_{V_{t,t-1}^{-1}}.$$

Applying the elliptical potential lemma per provider and using $\sum_i n_i = T - T_{\text{sub}}$ with Cauchy–Schwarz gives the stated $\tilde{O}(\sqrt{|T|dT})$ bound. $\square$

References

- [1] Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. 2011. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems*, Vol. 24. 2312–2320.
- [2] Mark Armstrong. 2006. Competition in two-sided markets. *The RAND journal of economics* 37, 3 (2006), 668–691.
- [3] Hemant K Bhargava. 2022. The creator economy: Managing ecosystem supply, revenue sharing, and platform design. *Management Science* 68, 7 (2022), 5233–5251.
- [4] Bernard Caillaud and Bruno Julien. 2003. Chicken & egg: Competition among intermediation service providers. *RAND journal of Economics* 34, 2 (2003), 309–328.
- [5] Alireza Chakeri and Luis G. Jaimes. 2018. An Incentive Mechanism for Crowdsensing Markets With Multiple Crowdsourcers. *IEEE Internet of Things Journal* 5, 2 (2018), 708–715.
- [6] David S Evans. 2016. Multisided platforms, dynamic competition, and the assessment of market power for internet-based firms. *University of Chicago Coase-Sandor Institute for Law & Economics Research Paper* 753 (2016).
- [7] David S Evans and Richard Schmalensee. 2013. *The antitrust analysis of multi-sided platform businesses*. Working Paper 18783. National Bureau of Economic Research.
- [8] Chongming Gao, Shichao Li, Wenqiang Lei, Jieming Chen, Ben Li, Peng Jiang, Xiangnan He, Jian Mao, and Tat-Seng Chua. 2022. KuaiRec: A fully-observed dataset and insights for evaluating recommender systems. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management*. 540–550.
- [9] Hengquan Guo, Ling kai Zu, and Xin Liu. 2025. Triple-Optimistic Learning for Stochastic Contextual Bandits with General Constraints. In *Proceedings of the 42nd International Conference on Machine Learning*. 21252–21276.
- [10] Kai Han, He Huang, and Jun Luo. 2018. Quality-Aware Pricing for Mobile Crowdsensing. *IEEE/ACM Transactions on Networking* 26, 4 (2018), 1728–1741.
- [11] Jiri Hron, Karl Krauth, Michael I. Jordan, Niki Kilbertus, and Sarah Dean. 2023. Modeling content creator incentives on algorithm-curated platforms. In *ICLR*. <https://openreview.net/forum?id=l6CpxxmUg>
- [12] Xinyan Hu, Meena Jagadeesan, Michael I Jordan, and Jacob Steinhardt. 2026. Incentivizing high-quality content in online recommender systems. In *Proceedings of the 7th Symposium on Foundations of Responsible Computing*. 15:1–15:18.
- [13] Nicole Immorlica, Meena Jagadeesan, and Brendan Lucier. 2024. Clickbait vs. quality: How engagement-based optimization shapes the content landscape in online platforms. In *Proceedings of the ACM Web Conference 2024*. 36–45.
- [14] Rolf Jagerman, Ilya Markov, and Maarten de Rijke. 2020. Safe Exploration for Optimizing Contextual Bandits. *ACM Transactions on Information Systems* 38, 3 (2020), 1–23.
- [15] Richard M Karp, Umesh V Vazirani, and Vijay V Vazirani. 1990. An optimal algorithm for on-line bipartite matching. In *Proceedings of the twenty-second annual ACM symposium on Theory of computing*. 352–358.
- [16] Haruka Kiyohara, Fan Yao, and Sarah Dean. 2025. Policy Design for Two-sided Platforms with Participation Dynamics. In *Proceedings of the 42nd International Conference on Machine Learning*. 30966–30985.
- [17] Simon Lacoste-Julien. 2016. Convergence rate of Frank-Wolfe for non-convex objectives. arXiv:1607.00345
- [18] Tor Lattimore and Csaba Szepesvári. 2020. *Bandit Algorithms*. Cambridge University Press.
- [19] Lihong Li, Wei Chu, John Langford, and Robert E Schapire. 2010. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*. 661–670.
- [20] Sige Liu, Peng Cheng, Zhuo Chen, Wei Xiang, Branka Vucetic, and Yonghui Li. 2023. Contextual User-Centric Task Offloading for Mobile Edge Computing in Ultra-Dense Network. *IEEE Transactions on Mobile Computing* 22, 9 (2023), 5092–5108.
- [21] Thodoris Lykouris, Vasilis Syrgkanis, and Eva Tardos. 2016. Learning and efficiency in games with dynamic population. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*. 120–129.
- [22] Aranyak Mehta, Amin Saberi, Umesh Vazirani, and Vijay Vazirani. 2007. Adwords and generalized online matching. *Journal of the ACM (JACM)* 54, 5 (2007), 22–es.
- [23] Meta for Business. 2025. Instagram Creator Marketplace. <https://www.facebook.com/business/ads/creator-marketplace>
- [24] Jer Shyuan Ng, Wei Yang Bryan Lim, Zehui Xiong, Dusit Niyato, Cyril Leung, and Chunyan Miao. 2022. A Double Auction Mechanism for Resource Allocation in Coded Vehicular Edge Computing. *IEEE Transactions on Vehicular Technology* 71, 2 (2022), 1832–1845.
- [25] Aldo Pacchiano, Mohammad Ghavamzadeh, and Peter Bartlett. 2025. Contextual bandits with stage-wise constraints. *Journal of machine learning research* 26, 170 (2025), 1–57.
- [26] Jean-Charles Rochet and Jean Tirole. 2003. Platform competition in two-sided markets. *Journal of the european economic association* 1, 4 (2003), 990–1029.
- [27] Jean-Charles Rochet and Jean Tirole. 2006. Two-sided markets: a progress report. *The RAND journal of economics* 37, 3 (2006), 645–667.
- [28] Sven Schmit and Carlos Riquelme. 2018. Human interaction with recommendation systems. In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*. 862–870.
- [29] Wen Sun, Jiajia Liu, Yanlin Yue, and Haibin Zhang. 2018. Double Auction-Based Resource Allocation for Mobile Edge Computing in Industrial Internet of Things. *IEEE Transactions on Industrial Informatics* 14, 10 (2018), 4692–4701.
- [30] TikTok. 2025. TikTok. <https://www.tiktok.com>
- [31] TikTok. 2025. TikTok One. <https://ads.tiktok.com/creative/creatormarketplace>
- [32] Huasen Wu, Rayadurgam Srikant, Xin Liu, and Chong Jiang. 2015. Algorithms with logarithmic or sublinear regret for constrained contextual bandits. In *Advances in Neural Information Processing Systems*, Vol. 28. 433–441.
- [33] Xiaohongshu. 2025. Xiaohongshu. <https://www.xiaohongshu.com>
- [34] Dejun Yang, Guoliang Xue, Xi Fang, and Jian Tang. 2016. Incentive Mechanisms for Crowdsensing: Crowdsourcing With Smartphones. *IEEE/ACM Transactions on Networking* 24, 3 (2016), 1732–1744.
- [35] Fan Yao, Yiming Liao, Jingzhou Liu, Shao liang Nie, Qifan Wang, Haifeng Xu, and Hongning Wang. 2024. Unveiling user satisfaction and creator productivity trade-offs in recommendation platforms. In *Advances in Neural Information Processing Systems*, Vol. 37. 86958–86984.
- [36] Fan Yao, Yiming Liao, Mingzhe Wu, Chuanhao Li, Yan Zhu, James Yang, Jingzhou Liu, Qifan Wang, Haifeng Xu, and Hongning Wang. 2024. User welfare optimization in recommender systems with competing content creators. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 3874–3885.
- [37] YouTube. 2025. BrandConnect for Influencer Advertising. <https://www.youtube.com/ads/brandconnect/>
- [38] Rongfei Zeng, Shixun Zhang, Jiaqi Wang, and Xiaowen Chu. 2020. FMore: An Incentive Scheme of Multi-dimensional Auction for Federated Learning in MEC. In *Proceedings of the 40th IEEE International Conference on Distributed Computing Systems*. 278–288.
- [39] Rui Zhang, Peng Cheng, Zhuo Chen, Sige Liu, Branka Vucetic, and Yonghui Li. 2022. Calibrated Bandit Learning for Decentralized Task Offloading in Ultra-Dense Networks. *IEEE Transactions on Communications* 70, 4 (2022), 2547–2560.
- [40] Xiao Zheng, Syed Bilal Hussian Shah, Saeeda Usman, Saoucene Mahfoudh, Fathima Shemim K S, and Piyush Kumar Shukla. 2023. Resource allocation and network pricing based on double auction in mobile edge computing. *Journal of Cloud Computing* 12, 1 (2023), 56.