REFINED REGRET ANALYSIS OF UCB-GS IN CENTRALIZED MATCHING MARKETS

Anonymous authors
Paper under double-blind review

ABSTRACT

In online two-sided matching markets, players seek to form stable matchings while learning their unknown preferences through repeated interactions with arms. A central platform can apply Gale–Shapley to players' upper confidence bound rankings of arms, an approach known as UCB-GS (Liu et al., 2020). We identify two sources of overcounting in the existing regret analysis of UCB-GS and sharpen it by jointly accounting for matchings blocked by the same blocking triplet and for different triplets that lead to observations of the same player-arm pair. Combining the first refinement with the structure of stable matchings, we obtain a player-pessimal stable regret bound of $O(K \log T/\Delta + N^2 \log T/\Delta^2)$ for each player over T rounds in a market with N players and K arms, where Δ is the minimum reward gap among each player's first $\min\{N+1, K\}$ arms. We also extend the analysis to many-to-one markets with responsive preferences. Two market families show that the two refinements can improve the preceding bounds by factors of order K and N, respectively, and that on each family the improved bound matches the lower bound up to constant factors when T is sufficiently large. Numerical experiments further illustrate these improvements.

1 INTRODUCTION

Two-sided matching markets arise in applications such as labor markets (Roth, 1984) and school admissions (Abdulkadiroğlu & Sönmez, 2003). In these markets, participants on each side have preferences over potential partners on the other side. A rich line of work studies stable matchings in such markets (Roth & Sotomayor, 1990). A matching is stable if no pair of participants on opposite sides would both prefer each other to their assigned partners. When all preferences are known, the classical Gale–Shapley (GS) algorithm finds a stable matching (Gale & Shapley, 1962).

In practice, however, participants on one side often do not know their preferences in advance. For example, on labor platforms such as Upwork and TaskRabbit, employers may not know how well workers will perform before assigning them tasks, while workers have their own preferences over employers (Liu et al., 2020). By repeatedly working with different workers and observing their performance, employers can gradually learn their preferences. Such learning from repeated feedback is naturally modeled by multi-armed bandits (Bubeck & Cesa-Bianchi, 2012; Lattimore & Szepesvári, 2020). In a bandit problem, a learner repeatedly selects an arm and observes a random reward only from the selected arm, so it must trade off exploring uncertain arms against exploiting arms that currently appear more rewarding. Classical bandit algorithms include explore-then-commit (ETC), upper confidence bound (UCB), and Thompson sampling (Lattimore & Szepesvári, 2020).

Liu et al. (2020) formalized bandit learning in matching markets by modeling participants with unknown preferences as players and participants on the other side as arms. The players' preferences over arms are determined by unknown mean rewards, which they learn from the rewards they observe when matched with an arm. As in prior work (Liu et al., 2020; Kong et al., 2024; 2025), arms have known preferences over players. The performance of each player is measured by stable regret, which compares its cumulative reward with that of a stable arm. Since a market may admit multiple stable matchings, the most and least preferred stable arms give rise to player-optimal and player-pessimal stable regret, respectively. Liu et al. (2020) considered a centralized setting in which a platform selects a matching in each round and studied both ETC and UCB approaches. The ETC policy uses an explicit exploration stage before committing to a matching. With prior knowledge of the horizon

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Table 1: Comparison of regret bounds for centralized matching markets.

	Assumptions	Requires T	Regret bound
Liu et al. (2020)	General	No	$O(NK \log T/\Delta_2^2)^{\#}$
Kong et al. (2024)	General	Yes	$O(N^2 \log T/\Delta_4^2 + K \log T/\Delta_4)^{*}$
Kong et al. (2024)	α-condition	No	$O(N \log T/\Delta_3^2 + K \log T/\Delta_4)^{*}$
This work	General	No	$O(N^2 \log T/\Delta^2 + K \log T/\Delta)^{\#}$
Wang et al. (2022)	General	No	$O(NK^3 \log(TN)/\Delta_2^2)^{\#}$
Kong & Li (2024)	General	Yes	$O(N \min\{N, K\} C \log T/\Delta_4^2)^{*}$
This work	General	No	$O(N \min\{N, K\} \log T/\Delta^2 + K \log T/\Delta)^{\#}$

The upper and lower parts correspond to one-to-one and many-to-one markets. All results assume strict preferences. $*$ and $\#$ denote player-optimal and player-pessimal stable regret, respectively, which coincide under the α-condition. In many-to-one markets, $C \geq N$ is the total capacity. Δ_2, $\Delta_3 = \Delta$, and Δ_4 denote the minimum reward gaps over all pairs of arms, over each player's first $\min\{N+1, K\}$ arms, and over each player's arms down to the one just below its player-optimal stable arm, respectively, and $\Delta_4 \geq \Delta_3 = \Delta \geq \Delta_2$.

T and the minimum reward gap, it achieves a player-optimal stable regret guarantee. The UCB policy, which we refer to as UCB-GS, instead recomputes the matching in every round by running Gale–Shapley on players' upper confidence bound rankings and arms' known preferences. UCB-GS has the practical advantage of not requiring the horizon T in advance, but in general admits only a player-pessimal stable regret guarantee.

For a market with N players and K arms, Liu et al. (2020) bound the player-pessimal stable regret of UCB-GS by $O(NK \log T/\Delta_2^2)$, where Δ_2 is the minimum reward gap over all players and all pairs of arms. For markets satisfying the α-condition, Kong et al. (2024) later refined the analysis of the same policy and removed the factor K from the leading term, which matters when there are many more arms than players. While this improvement exploits additional preference structure, it suggests that a more refined analysis of UCB-GS may be possible without such structure. This raises the question of how much the dependence on N and K in the existing regret bound can be improved in general markets.

To address this question, we revisit the regret analysis of UCB-GS in Liu et al. (2020). We identify two sources of overcounting. A single blocking triplet can block multiple matchings, while different blocking triplets can lead to observations of the same player-arm pair. Analyzing these matchings and triplets separately can therefore overcount their contributions to regret. Our main contributions are summarized as follows. (1) We develop two refinements that address the two sources of overcounting above. The first jointly accounts for all matchings blocked by the same triplet using one triplet count constraint, while the second derives joint constraints across triplets that increase the same player-arm observation count. (2) Building on the first refinement, we exploit the structure of stable matchings to establish a player-pessimal stable regret bound of $O(K \log T/\Delta + N^2 \log T/\Delta^2)$ for each player, where Δ is the minimum reward gap among each player's first $\min\{N+1, K\}$ arms. Compared with Liu et al. (2020), this removes the dependence on K from the main regret term and depends only on the reward gaps among these top-ranked arms. (3) For many-to-one markets with responsive preferences, where each arm can accept multiple players up to its capacity, we obtain a player-pessimal stable regret bound of $O(K \log T/\Delta + N \min\{N, K\} \log T/\Delta^2)$ for each player, sharpening the bound of Wang et al. (2022) for the same policy. We summarize the most closely related regret results in Table 1. (4) We construct two market families to characterize the effects of the two refinements. For the first family, the first refinement improves the bound of Liu et al. (2020) by a factor of order K when $T = \Omega(K^3 \log T)$. For the second family, the second refinement further improves the first refined bound by a factor of order N when $T = \Omega(K^2 \log T)$. We also establish corresponding regret lower bounds, showing that the improved bound captures the target player's regret under UCB-GS up to constant factors on each family for sufficiently large T.

2 Related Work

Bandit learning in matching markets dates back to Das & Kamenica (2005), who study a two-sided dating market. Liu et al. (2020) later introduce the competing bandits model with one-sided preference

uncertainty. They establish stable regret guarantees for centralized ETC and UCB-GS approaches, with the latter requiring no prior knowledge of T. Liu et al. (2021) and Kong et al. (2022) then propose decentralized CA-UCB and CA-TS, respectively, both with player-pessimal stable regret guarantees. Other studies relax the assumptions of strict preferences and one-sided uncertainty, allowing ties in preferences (Kong et al., 2025; Lin et al., 2025) or uncertainty on both sides of the market (Zhang & Fang, 2024; Basu, 2025). Another line of work studies the sample complexity of identifying a stable matching with high probability (Hosseini et al., 2024; Athanasopoulos et al., 2025; Pagare et al., 2025), and Li et al. (2025) provide a comprehensive survey of this area.

Several studies exploit preference structure to improve regret guarantees. Under serial dictatorship, where all arms rank players identically, Sankararaman et al. (2021) establish regret upper and lower bounds, and Wang & Li (2024) later match the lower bound. Basu et al. (2021) and Maheshwari et al. (2022) consider the broader α-condition and α-reducibility, respectively. Under the α-condition, Kong et al. (2024) sharpen the analysis of UCB-GS. For decentralized markets with arbitrary strict preferences, Basu et al. (2021) establish player-optimal guarantees through phased ETC, while Zhang et al. (2022) and Kong & Li (2023) develop ETC-type algorithms with improved guarantees. Kong et al. (2024) propose AETGS with player-optimal guarantees for both centralized and decentralized markets. We refine the regret analysis of UCB-GS for general markets with strict preferences.

In many-to-one markets, Wang et al. (2022) study bandit learning under responsive preferences in both centralized and decentralized settings. Further work considers preference conditions ensuring a unique stable matching (Guo et al., 2023) and quota constraints with complementary preferences (Li et al., 2024). Kong & Li (2024) further develop ETDA and AETDA with player-optimal stable regret guarantees, and Zhang & Fang (2026) develop SUBMARINE, which attains such guarantees in decentralized many-to-one markets. We extend our analysis to the centralized UCB policy of Wang et al. (2022) and obtain a sharper regret bound.

3 PRELIMINARIES

Problem setting. We consider a market with N players $\mathcal{P} = \{p_1, \dots, p_N\}$ and K arms $\mathcal{A} = \{a_1, \dots, a_K\}$, where a central platform coordinates their interactions. Each market participant has a preference ranking over potential partners on the other side. For player p_j, the reward from arm a_k follows a fixed distribution supported on $[0, 1]$ with unknown mean $\mu_{j,k}$. These means determine each player's preferences over arms, with $a_k \succ_j a_{k'}$ when $\mu_{j,k} > \mu_{j,k'}$. As in prior work (Liu et al., 2020; Basu et al., 2021; Kong et al., 2024), we assume $\mu_{j,k} \neq \mu_{j,k'}$ whenever $k \neq k'$, so each player has a strict preference ranking. Each arm a_k ranks the players according to a strict order π_k known to the platform, with $p_j \succ_{a_k} p_{j'}$ indicating that a_k prefers p_j to $p_{j'}$. We also assume $N \leq K$ so that every player can be matched with a distinct arm. Let $\mathcal{M}$ denote the set of all such matchings, where $m(j) = k$ means that p_j is matched with a_k in $m \in \mathcal{M}$.

At each round $t = 1, 2, \dots$, the platform selects a matching $m_t \in \mathcal{M}$, and each player p_j observes only its own reward $X_j(t)$ from its assigned arm $a_{m_t(j)}$, drawn independently from the corresponding reward distribution.

The players aim to form stable matchings under the true preferences $\succ_j$ and $\succ_{a_k}$. A pair (p_j, a_k) is a blocking pair for a matching m if player p_j prefers a_k to $a_{m(j)}$ and arm a_k is either unmatched or prefers p_j to its current partner. A matching is stable if it admits no blocking pair, but such a matching need not be unique. An arm is a stable arm of a player if the two are matched in some stable matching. The player-optimal and player-pessimal stable matchings assign each player its most and least preferred stable arms, respectively (Roth & Sotomayor, 1990). Let $\underline{m}$ denote the player-pessimal stable matching. We consider horizons $T \geq 2$. For a target player p_i, the player-pessimal stable regret over T rounds is $R_i(T) = \sum_{t=1}^{T} \mathbb{E}[\mu_{i,\underline{m}(i)} - \mu_{i,m_t(i)}]$, where the expectation is taken over the randomness of the rewards and the policy. This regret can be negative because a player may receive arms it prefers to its player-pessimal stable arm.

For a positive integer n, let $[n] = \{1, \dots, n\}$. For the subsequent regret analysis, we also define the following reward gap. The case $K = 1$ is trivial, so we assume $K \geq 2$ throughout.

Definition 1 (Minimum reward gap)**.** Let $\sigma_{j,r}$ denote the index of player p_j's r-th preferred arm under its true preferences. Define $\Delta = \min_{j \in [N]} \min_{1 \leq r \leq \min\{N, K-1\}} (\mu_{j,\sigma_{j,r}} - \mu_{j,\sigma_{j,r+1}})$ as the minimum reward gap among each player's top $\min\{N + 1, K\}$ arms.

UCB-GS. We now review the UCB-GS algorithm of Liu et al. (2020) (Algorithm 1). For $j \in [N]$ and $k \in [K]$, let $N_{j,k}(t)$ denote the number of rewards observed by player p_j from arm a_k up to round t. Let $\widehat{\mu}_{j,k}(t)$ denote the corresponding empirical mean. Define the UCB index as $U_{j,k}(t) = \widehat{\mu}_{j,k}(t-1) + \sqrt{3 \log t / (2N_{j,k}(t-1))}$ when $N_{j,k}(t-1) > 0$, and $U_{j,k}(t) = +\infty$ otherwise. At each round, players submit their UCB rankings to the platform. The platform then runs player-proposing GS based on the submitted rankings to match players and arms.

Algorithm 1 UCB-GS (Liu et al., 2020)

Require: N, K, and arm rankings $\{\pi_k\}_{k=1}^K$
1: $N_{j,k}(0) \leftarrow 0$, $\widehat{\mu}_{j,k}(0) \leftarrow 0$ for $j \in [N]$ and $k \in [K]$.
2: **for** $t = 1, 2, \ldots$ **do**
3: Each p_j submits a ranking $\widehat{\sigma}_j(t)$ of arms by decreasing $U_{j,k}(t)$, breaking ties arbitrarily.
4: $m_t \leftarrow \mathrm{GS}(\{\widehat{\sigma}_j(t)\}_{j=1}^N, \{\pi_k\}_{k=1}^K)$.
5: Each p_j observes $X_j(t)$ and updates $N_{j,k}(t)$ and $\widehat{\mu}_{j,k}(t)$ for $k = m_t(j)$.

Blocking triplets and the existing regret bound. We introduce the notation used in our analysis and recall the regret bound of Liu et al. (2020). For the target player p_i, define the regret gap $\Delta_{i,k} = \mu_{i,\underline{m}(i)} - \mu_{i,k}$ for each arm a_k. Every stable matching assigns p_i an arm weakly preferred to $a_{\underline{m}(i)}$, so any matching m with $\Delta_{i,m(i)} > 0$ is unstable under the true preferences and hence admits a blocking pair. A blocking triplet $(p_j, a_k, a_{k'})$ for m consists of a blocking pair (p_j, a_k) and the arm $a_{k'}$ assigned to p_j in m. Here, a_k is the blocking arm and $(p_j, a_{k'})$ is the assigned pair. Let $B_{j,k,k'}$ denote the set of matchings blocked by this triplet, and define the triplet gap as $\Delta_{j,k,k'} = \mu_{j,k} - \mu_{j,k'} > 0$. Let $T_m(T)$ denote the number of rounds in which matching m is selected up to round T. The triplet count $\sum_{m \in B_{j,k,k'}} T_m(T)$ is the number of rounds in which $(p_j, a_k, a_{k'})$ blocks the selected matching.

The analysis of Liu et al. (2020) derives upper bounds on the expected triplet counts and combines these bounds through covers. In each round t in which $(p_j, a_k, a_{k'})$ blocks the selected matching, player p_j observes a reward from $a_{k'}$. Since GS returns a matching stable under the submitted rankings, p_j must also rank $a_{k'}$ above a_k in $\widehat{\sigma}_j(t)$. Combining these two facts with concentration inequalities for the empirical means gives the following triplet count bound.

Lemma 2 (Liu et al. 2020)**.** *Under UCB-GS, for every blocking triplet $(p_j, a_k, a_{k'})$, we have $\sum_{m \in B_{j,k,k'}} \mathbb{E} T_m(T) \le 5 + 6 \log T / \Delta_{j,k,k'}^2$.*

To turn these bounds into a regret bound, write $R_i(T) = \sum_{\ell \in [K]} \Delta_{i,\ell} \, \mathbb{E} N_{i,\ell}(T)$, where arms with $\Delta_{i,\ell} \le 0$ contribute nonpositive terms. Let $\mathcal{M}_{i,\ell} \subseteq \mathcal{M}$ denote the set of unstable matchings that assign a_ℓ to p_i. A set Q of blocking triplets covers $\mathcal{M}_{i,\ell}$ if every matching in this set is blocked by at least one triplet in Q, and $\mathcal{C}(\mathcal{M}_{i,\ell})$ denotes the collection of all such covers. For each arm a_ℓ with $\Delta_{i,\ell} > 0$, every round in which p_i receives a_ℓ selects a matching in $\mathcal{M}_{i,\ell}$, so $N_{i,\ell}(T)$ is at most the sum of the triplet counts over any cover of $\mathcal{M}_{i,\ell}$. Applying Lemma 2 and minimizing over covers gives $R_i(T) \le \Lambda_i(T)$ (Liu et al., 2020, Theorem 3), where

$$\Lambda_i(T) = \sum_{\ell : \Delta_{i,\ell} > 0} \Delta_{i,\ell} \min_{Q \in \mathcal{C}(\mathcal{M}_{i,\ell})} \sum_{(p_j, a_k, a_{k'}) \in Q} \left(5 + \frac{6 \log T}{\Delta_{j,k,k'}^2} \right). \tag{1}$$

4 Refined Regret Analysis

We refine $\Lambda_i(T)$ by sharing triplet counts across matchings (Section 4.1) and observation counts across triplets (Section 4.2).

4.1 Sharing Triplet Counts Across Matchings

A triplet can block matchings that assign different arms to the target player, so its triplet count bound may appear in the terms of (1) corresponding to these arms. For example, consider two players with rankings $a_1 \succ_1 a_2 \succ_1 a_3$ and $a_2 \succ_2 a_1 \succ_2 a_3$, and three arms that all rank $p_1 \succ p_2$. With p_2 as the target player, the matchings $m = \{(p_1, a_2), (p_2, a_1)\}$ and $m' = \{(p_1, a_2), (p_2, a_3)\}$ are both

blocked by (p_1, a_1, a_2). Since $m \in \mathcal{M}_{2,1}$ and $m' \in \mathcal{M}_{2,3}$, the triplet count bound of (p_1, a_1, a_2) may appear in both the $\ell = 1$ and $\ell = 3$ terms of (1). However, since $B_{1,1,2} = \{m, m'\}$, this bound applies to $\mathbb{E}T_m(T) + \mathbb{E}T_{m'}(T)$ jointly.

In general, each triplet count bound controls the sum of $\mathbb{E}T_m(T)$ over all matchings blocked by that triplet, so we impose it as a single constraint on this sum. We first write the regret in terms of the same counts as $R_i(T) = \sum_{m \in \mathcal{M}} \Delta_{i,m(i)} \, \mathbb{E}T_m(T)$. We then upper bound it by replacing each $\mathbb{E}T_m(T)$ with a variable $x_m \geq 0$ and maximizing subject to these triplet constraints and the horizon constraint, which gives

$$\begin{aligned} \Gamma_i(T) = \max_{x_m \geq 0} \quad & \sum_{m \in \mathcal{M}} \Delta_{i,m(i)} x_m \\ \text{subject to} \quad & \sum_{m \in B_{j,k,k'}} x_m \leq 5 + \frac{6 \log T}{\Delta^2_{j,k,k'}} \quad \text{for every blocking triplet } (p_j, a_k, a_{k'}), \\ & \sum_{m \in \mathcal{M}} x_m \leq T. \end{aligned} \tag{2}$$

The following theorem shows that $\Gamma_i(T)$ upper bounds the regret and never exceeds $\Lambda_i(T)$.

Theorem 3. *For every target player p_i, we have $R_i(T) \leq \Gamma_i(T) \leq \Lambda_i(T)$.*

Taking $x_m = \mathbb{E}T_m(T)$ gives a feasible solution with objective value $R_i(T)$, so $R_i(T) \leq \Gamma_i(T)$. For any feasible solution and any arm a_ℓ with $\Delta_{i,\ell} > 0$, summing the triplet constraints over a cover of $\mathcal{M}_{i,\ell}$ bounds $\sum_{m \in \mathcal{M}_{i,\ell}} x_m$. Discarding nonpositive terms and grouping the rest by the arm assigned to p_i bounds the objective by $\sum_{\ell : \Delta_{i,\ell} > 0} \Delta_{i,\ell} \sum_{m \in \mathcal{M}_{i,\ell}} x_m$. Minimizing this count bound over covers of each $\mathcal{M}_{i,\ell}$ shows that this sum is at most $\Lambda_i(T)$. Appendix D.1 gives the full proof.

4.2 Sharing Observation Counts Across Triplets

The linear program for $\Gamma_i(T)$ still imposes a separate count constraint for each blocking triplet, even when several triplets share the same assigned pair. For example, consider two players with rankings $a_1 \succ_1 a_2 \succ_1 a_3$ and $a_2 \succ_2 a_1 \succ_2 a_3$. Arms a_1 and a_2 rank $p_2 \succ p_1$, while a_3 ranks $p_1 \succ p_2$. With p_1 as the target player, consider the matchings $m = \{(p_1, a_3), (p_2, a_2)\}$ and $m' = \{(p_1, a_3), (p_2, a_1)\}$. The matching m is blocked only by (p_1, a_1, a_3), while m' is blocked only by (p_1, a_2, a_3) and (p_2, a_2, a_1). These matchings have no blocking triplet in common, so their counts do not appear together in any triplet constraint. However, both matchings assign a_3 to p_1, so every round that selects either matching increases the same observation count $N_{1,3}$.

To exploit this shared observation count, fix an assigned pair $(p_j, a_{k'})$. Every round in which some triplet $(p_j, a_k, a_{k'})$ blocks the selected matching increases $N_{j,k'}$, whichever arm a_k forms the blocking pair. To bound the number of these rounds, we use the following consequence of stability under the submitted rankings.

Lemma 4 (Liu et al., 2020)**.** *Suppose player p_j receives arm $a_{k'}$ in round t. If (p_j, a_k) blocks m_t under the true preferences, then $U_{j,k'}(t) \geq U_{j,k}(t)$.*

Different blocking arms may have different triplet gaps. We therefore fix a threshold $d > 0$ and consider rounds in which p_j receives $a_{k'}$ and forms a blocking pair with some arm a_k satisfying $\Delta_{j,k,k'} \geq d$. Let $\mathcal{S}_{j,k'}(d) = \bigcup_{k : \Delta_{j,k,k'} \geq d,\, B_{j,k,k'} \neq \varnothing} B_{j,k,k'}$ denote the set of matchings satisfying this condition, and let $H_{j,k'}(d) = |\{k : \Delta_{j,k,k'} \geq d,\ B_{j,k,k'} \neq \varnothing\}|$ be the number of corresponding blocking arms. The sum $\sum_{m \in \mathcal{S}_{j,k'}(d)} T_m(T)$ counts each such round once, even if multiple blocking arms satisfy the threshold. The following theorem bounds the expectation of this sum using Lemma 4.

Theorem 5. *For every $j \in [N]$, $k' \in [K]$, and $d > 0$, we have*

$$\sum_{m \in \mathcal{S}_{j,k'}(d)} \mathbb{E}T_m(T) \leq C_{j,k'}(d, T) := \min \left\{ T,\ \left\lfloor \frac{6 \log T}{d^2} \right\rfloor + 2 + 2\sqrt{H_{j,k'}(d) + 1} \right\}. \tag{3}$$

The $\log T$ term avoids the factor $H_{j,k'}(d)$ from summing separate triplet count bounds using the same threshold d. Once $N_{j,k'}(t-1) > 6 \log T / d^2$, each such round requires a confidence failure

by Lemma 4. Balancing earlier rounds against later failures over the assigned arm and the arms in $\mathcal{A}_{j,k'}(d)$ yields the term $2\sqrt{H_{j,k'}(d)+1}$. Appendix B gives the proof.

To incorporate this count bound into the regret analysis, let $\Psi_i(T)$ denote the optimal value of (2) with the joint constraints $\sum_{m\in\mathcal{S}_{j,k'}(d)} x_m \le C_{j,k'}(d,T)$ from Theorem 5 for every $j \in [N]$, $k' \in [K]$, and $d \in \{\Delta_{j,k,k'} : k \in [K],\ B_{j,k,k'} \neq \varnothing\}$. Appendix C gives the full linear program and shows that only thresholds equal to relevant triplet gaps are needed. These joint constraints yield the following refinement of Theorem 3.

Theorem 6. *For every target player p_i, $R_i(T) \le \Psi_i(T) \le \Gamma_i(T) \le \Lambda_i(T)$.*

Taking $x_m = \mathbb{E}T_m(T)$ satisfies the triplet and horizon constraints and the joint constraints from Theorem 5, with objective value $R_i(T)$, so $R_i(T) \le \Psi_i(T)$. Removing the joint constraints recovers (2), so $\Psi_i(T) \le \Gamma_i(T)$. The last inequality follows from Theorem 3. We next quantify how much $\Psi_i(T)$ can improve on $\Gamma_i(T)$.

Proposition 7. *For every target player p_i, we have $\Psi_i(T) \le \Gamma_i(T) \le 2N\Psi_i(T)$.*

The proof uses the overlap among the triplet sets in each union $\mathcal{S}_{j,k'}(d)$ to show that dividing any feasible solution of (2) by $2N$ gives a feasible solution for $\Psi_i(T)$ (Appendix D.2). Section 6.2 shows that the factor N in this comparison is tight up to a constant factor.

5 Regret Guarantees

5.1 One-to-One Markets

We first establish a regret bound for UCB-GS in one-to-one markets with arbitrary strict preferences.

Theorem 8. *For every target player p_i, $R_i(T) \le \Gamma_i(T) \le O(K\log T/\Delta + N^2\log T/\Delta^2)$.*

We prove Theorem 8 by applying Lemma 2 to a cover based on the structure of stable matchings. To motivate this construction, we first examine a direct cover of the matchings with positive regret gaps. These matchings assign p_i an arm worse than its player-pessimal stable arm and are therefore unstable. By Lemma 20, they can be covered by blocking triplets $(p_j, a_k, a_{k'})$ with a_k among p_j's first N choices. For each assigned pair $(p_j, a_{k'})$ in this cover, the r-th smallest gap among the corresponding triplets is at least $r\Delta$, so their triplet count bounds sum to $O(\log T/\Delta^2)$ (Lemma 22). Since this cover involves at most $N(K-1)$ assigned pairs and each positive regret gap is at most one, summing over these pairs gives a regret bound of order $NK\log T/\Delta^2$.

To improve this dependence on K, we use the set $S = \{a_{\underline{m}(j)} : j \in [N]\}$ of arms assigned by the player-pessimal stable matching. Since S contains only N arms, the matchings in which a player assigned to an arm in S forms a blocking pair can be covered using at most N^2 assigned pairs. After separately handling matchings blocked by $(p_i, a_{\underline{m}(i)})$, we use the stability of $\underline{m}$ to cover the remaining matchings with few additional triplets. The following proposition partitions the matchings with positive regret gaps into three classes.

Proposition 9. *For every target player p_i, the matchings with positive regret gaps fall into three classes. (1) The pair $(p_i, a_{\underline{m}(i)})$ blocks the matching. (2) The matching is outside the first class, but some player assigned to an arm in S forms a blocking pair. Any such player also forms a blocking pair with an arm among its first N choices. (3) The matching belongs to neither of the first two classes. If this class is nonempty, a common arm outside S is assigned in all its matchings. In each such matching, the player holding this arm forms a blocking pair with its player-pessimal stable arm.*

If the third class is nonempty, we construct a stable matching in the market restricted to S for each matching in this class, preserving the target's assignment, and take their player join (Roth & Sotomayor, 1990). This join is stable on S but still assigns the target an arm worse than its player-pessimal stable arm $a_{\underline{m}(i)}$, so its extension to the full market admits a blocking pair with an arm outside S. The join property, together with the stability of $\underline{m}$ and the defining conditions of the third class, implies that this arm is assigned in every matching of that class (Appendix F).

We now bound the contribution of each class. (1) In the first class, if p_i receives a_ℓ, the matching is blocked by the triplet $(p_i, a_{\underline{m}(i)}, a_\ell)$, whose triplet gap $\Delta_{i,\underline{m}(i),\ell}$ equals $\Delta_{i,\ell}$. By Lemma 19, $a_{\underline{m}(i)}$

is among p_i's first N choices. Since p_i prefers this arm to a_ℓ, Definition 1 implies $\Delta_{i,\ell} \geq \Delta$. By Lemma 2, each a_ℓ contributes $\Delta_{i,\ell} \cdot O(\log T/\Delta_{i,\ell}^2) = O(\log T/\Delta)$. Summing over at most $K-1$ arms gives $O(K \log T/\Delta)$. (2) Each matching in the second class is blocked by a triplet $(p_j, a_k, a_{k'})$ with $a_{k'} \in S$ and a_k among p_j's first N choices. These triplets involve at most N^2 assigned pairs, each contributing $O(\log T/\Delta^2)$ by Lemma 22, for a total of $O(N^2 \log T/\Delta^2)$. (3) The holder of the common arm may vary across matchings in the third class, and each possible holder gives one triplet, so at most N triplets cover this class. As in the first class, their gaps are at least Δ, so Lemma 2 gives a total contribution of $O(N \log T/\Delta^2)$, which is absorbed into the second term.

These estimates bound $\Gamma_i(T)$ using only the triplet constraints in (2). The regret bound then follows from Theorem 3. Appendix F.3 gives the full proof. Appendix F.5 gives a finer bound using a separate minimum reward gap for each player.

5.2 Many-to-One Markets

We now consider many-to-one markets with responsive preferences. Each arm a_k has capacity $c_k \geq 1$ and a strict ranking of players known to the platform. Let $\mathcal{M}_c$ denote the set of matchings that assign one arm to each player and at most c_k players to each arm a_k. We assume $\sum_{k=1}^{K} c_k \geq N$, so $\mathcal{M}_c$ is nonempty even when $K < N$. A pair blocks a matching if the player prefers the arm to its assignment and the arm either has a vacancy or prefers the player to one of its current holders. Rewards follow the model in Section 3, and regret is measured relative to this market's player-pessimal stable matching. We retain the gap Δ from Definition 1. The platform can still use the UCB policy, with each arm keeping up to c_k highest-ranked proposing players during player-proposing GS (Wang et al., 2022).

We define a seat market by replacing each arm a_k with c_k seats that share its ranking of players. Each player orders seats by its original arm preferences and prefers smaller seat indices within each arm. Since player-proposing GS with capacities still returns a matching stable under the submitted rankings, Lemma 2 and the linear programs of Section 4 carry over to many-to-one markets. We then construct covers on the seat market to obtain the following regret bounds.

Theorem 10. *For every target player p_i, $R_i(T) \leq O(K \log T/\Delta + N \min\{N, K\} \log T/\Delta^2)$. If the seat market satisfies the α-condition defined in Appendix J.5, then $R_i(T) \leq O(K \log T/\Delta + N \log T/\Delta^2)$.*

Since the seat market is one-to-one and its stable matchings correspond to those of the many-to-one market (Roth & Sotomayor, 1989), we apply the structural argument of Section 5.1 to it. Assigning each arm's holders to its first seats in the order of that arm's ranking rules out blocking pairs between a player and another seat of its assigned arm, so replacing each seat by its arm yields a cover by triplets with positive gaps. The player-pessimal stable matching uses at most $\min\{N, K\}$ original arms, so the second class involves at most $N \min\{N, K\}$ assigned pairs. Under the α-condition, the cover uses each player's stable arm as the blocking arm and, for players other than the target, another arm used by $\underline{m}$ as the assigned arm. Summing the corresponding triplet count bounds gives $O(\log T/\Delta^2)$ for each such player, yielding the $N \log T/\Delta^2$ term. Appendix J gives the full proof.

When $c_k = 1$ for all $k \in [K]$, the market becomes one-to-one and coincides with its seat market. The α-condition on the seat market then reduces to the one used by Kong et al. (2024), and the second bound has the same dependence on N and K as theirs.

6 Tightness on Two Market Families

We construct two market families to quantify the improvements from the two refinements. On each family, we show that the improved bound captures the target player's regret under UCB-GS up to constant factors for sufficiently large T.

6.1 Triplet Count Sharing

We consider a market M_K^{A} with $N = 2$ players and $K \geq 4$ arms, and take p_2 as the target player. Every arm ranks $p_1 \succ p_2$. We take independent Bernoulli rewards with means $\mu_{1,1} = \mu_{2,2} = 1/2$, $\mu_{1,2} = 1/2 - 1/K$, and $\mu_{j,\ell} = 1/(K\ell)$ for all remaining player-arm pairs. Thus p_1 ranks $a_1 \succ a_2 \succ$

$\cdots \succ a_K$, while p_2 ranks $a_2 \succ a_1 \succ a_3 \succ \cdots \succ a_K$. The unique stable matching therefore assigns a_1 to p_1 and a_2 to p_2. Relative to this matching, assigning any a_ℓ with $\ell \neq 2$ to p_2 gives regret gap $\Delta_{2,\ell} = 1/2 - 1/(K\ell) \in [1/4, 1/2)$, whereas the triplet (p_1, a_1, a_2) has gap $\Delta_{1,1,2} = 1/K$. We now compare the regret bounds $\Lambda_2(T)$, $\Gamma_2(T)$, and $\Psi_2(T)$ on this market.

Theorem 11. *On the market M_K^{A}, $\Lambda_2(T) = \Theta(K^3 \log T)$, whereas $\Gamma_2(T) = \Theta(\min\{T, K^2 \log T\})$ and $\Psi_2(T) = \Theta(\min\{T, K^2 \log T\})$. In particular, if $T \geq K^3 \log T$, then $\Lambda_2(T)/\Gamma_2(T) = \Theta(K)$.*

The separation comes from the matchings $m_\ell = \{(p_1, a_2), (p_2, a_\ell)\} \in \mathcal{M}_{2,\ell}$ with $\ell \neq 2$, which share the blocking triplet (p_1, a_1, a_2). Because $\Lambda_2(T)$ uses a separate cover for each $\mathcal{M}_{2,\ell}$, each cover must contain a triplet blocking m_ℓ, and every such triplet has gap at most $1/K$. Using $\Delta_{2,\ell} \geq 1/4$ in (1), we obtain $\Lambda_2(T) = \Omega(K^3 \log T)$. In contrast, in the linear program for $\Gamma_2(T)$, the triplet constraint for (p_1, a_1, a_2) alone bounds the total count of these matchings by $O(K^2 \log T)$. Appendix G gives the full proof.

We next establish a regret lower bound on M_K^{A}. Let $R_i^\pi(T; M)$ denote player p_i's regret under a learning algorithm π in market M, and continue to write $R_i(T)$ for the regret of UCB-GS.

Theorem 12. *Let $M_K^{\mathrm{A}\prime}$ be obtained from M_K^{A} by setting $\mu_{1,2} = 1/2 + 1/K$. For any learning algorithm π satisfying $R_1^\pi(T; M_K^{\mathrm{A}}) = o(T^\alpha)$ and $R_1^\pi(T; M_K^{\mathrm{A}\prime}) = o(T^\alpha)$ for every $\alpha > 0$, we have*

$$\liminf_{T\to\infty} \frac{R_2^\pi(T; M_K^{\mathrm{A}})}{\log T} \geq \frac{K^2}{48}.$$

Moreover, on M_K^{A}, if $T \geq 64K^2 \log T$, then $R_2(T) = \Theta(K^2 \log T) = \Theta(\Gamma_2(T)) = \Theta(\Psi_2(T))$.

The lower bound arises because M_K^{A} and $M_K^{\mathrm{A}\prime}$ differ only in $\mu_{1,2}$ by $2/K$. Any algorithm with subpolynomial regret for p_1 in both markets must therefore match p_1 with a_2 $\Omega(K^2 \log T)$ times in expectation to distinguish them. In each such round, p_2 cannot receive a_2 and incurs regret at least $1/2 - 1/K$, giving a regret lower bound of $\Omega(K^2 \log T)$. A similar phenomenon appears in Liu et al. (2020, Example 2). For UCB-GS, the upper bound $R_2(T) \leq \Gamma_2(T) = O(K^2 \log T)$ matches this lower bound, so $\Gamma_2(T)$ is tight. Appendix I.2 gives the full proof. Since $N = 2$ and $\Delta = 2/(3K)$, the general bound in Theorem 8 reduces to $O(K^2 \log T)$ and is also tight on this market.

6.2 Observation Count Sharing

We next consider a market M_K^{B} with $N = K - 1$ players and $K \geq 3$ arms, and take p_1 as the target player. Every arm a_k with $k < K$ ranks $p_2 \succ \cdots \succ p_{K-1} \succ p_1$, while a_K ranks $p_1 \succ p_2 \succ \cdots \succ p_{K-1}$. Let $\zeta = 1/(4\sqrt{K})$. We take independent Bernoulli rewards with means $\mu_{1,K} = 1/2$, $\mu_{1,k} = 1/2 + \zeta(1 - (k-1)/K^3)$ for $k < K$, and, for each $2 \leq j < K$, $\mu_{j,j} = 1/2 + \zeta$ and $\mu_{j,k} = 1/2 + \zeta/(k+1)$ for $k \neq j$. Thus p_1 ranks $a_1 \succ a_2 \succ \cdots \succ a_K$, while every other player ranks its own arm first and the remaining arms by increasing index. The unique stable matching assigns a_j to p_j for every $j < K$ and leaves a_K unmatched. Relative to this matching, assigning a_K to p_1 gives regret gap $\Delta_{1,K} = \zeta$, and every triplet (p_1, a_j, a_K) with $j < K$ has gap $\Delta_{1,j,K} \in [\zeta/2, \zeta]$. We compare the three regret bounds on this market and then establish a regret lower bound on M_K^{B}.

Theorem 13. *On M_K^{B}, $\Lambda_1(T) = \Theta(K^{3/2} \log T)$, $\Gamma_1(T) = \Theta(\min\{T/\sqrt{K}, K^{3/2} \log T\})$, and $\Psi_1(T) = \Theta(\min\{T/\sqrt{K}, \sqrt{K} \log T\})$. In particular, if $T \geq K^2 \log T$, then $\Gamma_1(T)/\Psi_1(T) = \Theta(N)$ and $\Lambda_1(T)/\Psi_1(T) = \Theta(N)$, where $N = K - 1$.*

The separation comes from $K - 1$ matchings that all assign a_K to p_1 but have pairwise disjoint sets of blocking triplets. Because these sets are disjoint and their triplets have gaps at most ζ, the linear program for $\Gamma_1(T)$ can assign each matching a count of order $\log T/\zeta^2$ when $T \geq K^2 \log T$, so $\Gamma_1(T) = \Omega(K \log T/\zeta)$. Since $\Lambda_1(T)$ has the same order in this regime, triplet count sharing does not improve the order of the bound. However, every round selecting one of these matchings increases the same observation count $N_{1,K}$, and the unmatched arm gives a blocking triplet with gap at least $\zeta/2$. The joint constraint from Theorem 5 therefore bounds these rounds by $O(\log T/\zeta^2)$, so they contribute only $O(\log T/\zeta)$ to $\Psi_1(T)$. Thus Proposition 7 is tight in its dependence on N. Appendix H gives the full proof.

Theorem 14. *Let $M_K^{\mathrm{B}\prime}$ be obtained from M_K^{B} by setting $\mu_{1,K} = 1/2+2\zeta$. For any learning algorithm π satisfying $R_1^\pi(T; M_K^{\mathrm{B}}) = o(T^\alpha)$ and $R_1^\pi(T; M_K^{\mathrm{B}\prime}) = o(T^\alpha)$ for every $\alpha > 0$, we have*

$$\liminf_{T\to\infty} \frac{R_1^\pi(T; M_K^{\mathrm{B}})}{\log T} \geq \frac{\sqrt{K}}{3}.$$

Moreover, on M_K^{B}, if $T \geq 384K^2 \log T$, then $R_1(T) = \Theta(\sqrt{K}\log T) = \Theta(\Psi_1(T))$.

As in Section 6.1, any algorithm with subpolynomial regret for p_1 in both markets needs $\Omega(\log T/\zeta^2)$ observations of (p_1, a_K) in expectation to distinguish them. Each such observation incurs regret ζ for p_1, giving a regret lower bound of $\Omega(\sqrt{K}\log T)$. For UCB-GS, the upper bound $R_1(T) \leq \Psi_1(T) = O(\sqrt{K}\log T)$ matches this lower bound, so $\Psi_1(T)$ is tight. Appendix I.3 gives the full proof.

7 Experiments

We compare $\Lambda_1(t)$, $\Gamma_1(t)$, and $\Psi_1(t)$ with the empirical player-pessimal stable regret of UCB-GS on random markets. Both sides' preference rankings are sampled independently as uniform permutations. We fix $K = 10$ and vary $N \in \{2, 3, 4, 5\}$. For every player, the mean rewards decrease from 0.95 to 0.05 in steps of 0.1 along its ranking. Rewards are independent Bernoulli variables with these means. For each value of N, we generate 20 markets and run UCB-GS for $T = 100{,}000$ rounds over 50 independent repetitions per market. We report the regret of p_1, averaged over repetitions and markets. Appendix K gives additional results and computational details, including the cap on $\Lambda_1(t)$.

Figure 1 shows that both triplet count sharing and observation count sharing reduce the mean regret bound across all four settings. At $T = 100{,}000$, the mean $\Gamma_1(T)$ is about 6%, 14%, 30%, and 26% lower than the mean $\Lambda_1(T)$ for $N = 2, 3, 4, 5$, respectively. The mean $\Psi_1(T)$ is about 0.5% to 4.7% lower than the mean $\Gamma_1(T)$. For $N = 4$, the mean regret becomes negative because p_1's gains from receiving arms preferred to its player-pessimal stable arm outweigh its losses from worse arms.

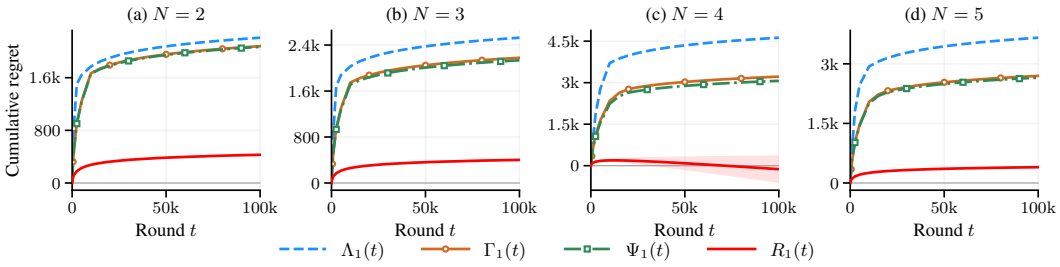

Figure 1: Regret bounds and empirical player-pessimal stable regret of UCB-GS. Shading shows one standard error of the market averages.

8 Conclusion

We revisited the regret analysis of UCB-GS and addressed two sources of overcounting by jointly accounting for matchings blocked by the same triplet and for triplets that share observations of the same player-arm pair. Combined with the structure of stable matchings, the first refinement yielded a player-pessimal stable regret bound of $O(K\log T/\Delta + N^2\log T/\Delta^2)$, which removes the dependence on K in the main order term of the bound of Liu et al. (2020). We also extended the analysis to many-to-one markets. Two market families show that the corresponding refined bounds capture the target players' regret under UCB-GS up to constant factors for sufficiently large T, while experiments on random markets further illustrate the improvements.

One future direction is to determine whether the quadratic dependence on N is necessary in general markets. The known lower bound, established under serial dictatorship (Sankararaman et al., 2021), grows only linearly in N. Our analysis uses player-proposing GS only through the stability of its output under the submitted rankings, so exploiting further properties of GS may improve the dependence on N.

STATEMENTS

AI USE STATEMENT

The authors used generative AI tools as an interactive aid in implementing the experimental code, checking and refining parts of the proofs, searching for and comparing related work, and improving the clarity of the manuscript. All mathematical arguments and experimental code involving AI assistance, as well as the reported results, references, and comparisons with related work, were reviewed and verified by the authors. The authors take full responsibility for the final content of this paper.

ETHICS STATEMENT

This work studies regret guarantees for learning in two-sided matching markets. All experiments use synthetic markets with rewards sampled from Bernoulli distributions. No human participants or personal data are involved.

REPRODUCIBILITY STATEMENT

The model assumptions and UCB policy are specified in Section 3. Complete proofs of the theoretical results are provided in Appendices A–J. Section 7 and Appendix K describe the experimental setup, market generation, and computation of the regret bounds.

REFERENCES

Atila Abdulkadiroğlu and Tayfun Sönmez. School choice: A mechanism design approach. *American economic review*, 93(3):729–747, 2003.

Andreas Athanasopoulos, Anne-Marie George, and Christos Dimitrakakis. Probably correct optimal stable matching for two-sided market under uncertainty. In *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems*, AAMAS '25, pp. 152–160. International Foundation for Autonomous Agents and Multiagent Systems, 2025. ISBN 9798400714269.

Soumya Basu. Competing bandits in matching markets via super stability. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 3226–3250. PMLR, 13–19 Jul 2025.

Soumya Basu, Karthik Abinav Sankararaman, and Abishek Sankararaman. Beyond $\log^2(T)$ regret for decentralized bandits in matching markets. In *International Conference on Machine Learning*, pp. 705–715. PMLR, 2021.

Yosef Blum, Alvin E Roth, and Uriel G Rothblum. Vacancy chains and equilibration in senior-level labor markets. *Journal of Economic theory*, 76(2):362–411, 1997.

Sébastien Bubeck and Nicolò Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5(1):1–122, 2012.

Vasek Chvatal. A greedy heuristic for the set-covering problem. *Mathematics of operations research*, 4(3):233–235, 1979.

Sanmay Das and Emir Kamenica. Two-sided bandits and the dating market. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, pp. 947–952, 2005.

David Gale and Lloyd S Shapley. College admissions and the stability of marriage. *The American mathematical monthly*, 69(1):9–15, 1962.

L. Guo, Z. Wang, and S. Li. Bandit learning in many-to-one matching markets with uniqueness conditions. OpenReview, 2023. URL `https://openreview.net/forum?id=VgCqZL_N0a`.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

Hadi Hosseini, Sanjukta Roy, and Duohan Zhang. Putting Gale & Shapley to work: Guaranteeing stability through learning. In *Advances in Neural Information Processing Systems*, volume 37, pp. 69043–69068. Curran Associates, Inc., 2024.

Fang Kong and Shuai Li. Player-optimal stable regret for bandit learning in matching markets. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 1512–1522. SIAM, 2023.

Fang Kong and Shuai Li. Improved bandits in many-to-one matching markets with incentive compatibility. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 13256–13264, 2024.

Fang Kong, Junming Yin, and Shuai Li. Thompson sampling for bandit learning in matching markets. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pp. 3164–3170. International Joint Conferences on Artificial Intelligence Organization, 7 2022.

Fang Kong, Zilong Wang, and Shuai Li. Improved analysis for bandit learning in matching markets. In *Proceedings of the 38th International Conference on Neural Information Processing Systems*, volume 37, pp. 91904–91929, 2024.

Fang Kong, Jingqi Tang, Mingzhu Li, Pinyan Lu, John C.S. Lui, and Shuai Li. Bandit learning in matching markets with indifference. In *International Conference on Learning Representations*, volume 2025, pp. 23627–23645, 2025.

Tor Lattimore and Csaba Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020.

Shuai Li, Zilong Wang, and Fang Kong. A survey on bandit learning in matching markets. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, pp. 10546–10554. International Joint Conferences on Artificial Intelligence Organization, 8 2025.

Yuantong Li, Guang Cheng, and Xiaowu Dai. Two-sided competing matching recommendation markets with quota and complementary preferences constraints. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pp. 27589–27614, 2024.

Shiyun Lin, Simon Mauras, Nadav Merlis, and Vianney Perchet. Stable matching with ties: Approximation ratios and learning. In *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 96661–96699, 2025.

Lydia T Liu, Horia Mania, and Michael I. Jordan. Competing bandits in matching markets. In *International Conference on Artificial Intelligence and Statistics*, pp. 1618–1628. PMLR, 2020.

Lydia T Liu, Feng Ruan, Horia Mania, and Michael I. Jordan. Bandit learning in decentralized matching markets. *Journal of Machine Learning Research*, 22(211):1–34, 2021.

Chinmay Maheshwari, Shankar Sastry, and Eric Mazumdar. Decentralized, communication- and coordination-free learning in structured matching markets. In *Advances in Neural Information Processing Systems*, volume 35, pp. 15081–15092. Curran Associates, Inc., 2022.

Tejas Pagare, Agniv Bandyopadhyay, and Sandeep Juneja. Optimal algorithms for bandit learning in matching markets. *arXiv preprint arXiv:2509.14466*, 2025.

A.E. Roth and M.A.O. Sotomayor. *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Econometric Society Monographs. Cambridge University Press, 1990. ISBN 9780521437882.

Alvin E Roth. The evolution of the labor market for medical interns and residents: a case study in game theory. *Journal of political Economy*, 92(6):991–1016, 1984.

Alvin E Roth and Marilda Sotomayor. The college admissions problem revisited. *Econometrica: Journal of the Econometric Society*, pp. 559–570, 1989.

Abishek Sankararaman, Soumya Basu, and Karthik Abinav Sankararaman. Dominate or delete: Decentralized competing bandits in serial dictatorship. In *International Conference on Artificial Intelligence and Statistics*, pp. 1252–1260. PMLR, 2021.

Zilong Wang and Shuai Li. Optimal analysis for bandit learning in matching markets with serial dictatorship. *Theoretical Computer Science*, 1010:114703, 2024.

Zilong Wang, Liya Guo, Junming Yin, and Shuai Li. Bandit learning in many-to-one matching markets. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pp. 2088–2097, 2022.

Yirui Zhang and Zhixuan Fang. Decentralized two-sided bandit learning in matching market. In *Proceedings of the Fortieth Conference on Uncertainty in Artificial Intelligence*, volume 244 of *Proceedings of Machine Learning Research*, pp. 4173–4191. PMLR, 15–19 Jul 2024.

Yirui Zhang and Zhixuan Fang. Decentralized many-to-one resource allocation: The competing bandit approach. *IEEE Transactions on Networking*, 34:2149–2164, 2026.

YiRui Zhang, Siwei Wang, and Zhixuan Fang. Matching in multi-arm bandit with collision. In *Advances in Neural Information Processing Systems*, volume 35, pp. 9552–9563, 2022.

A Proofs of Lemmas 2 and 4

For completeness, we reproduce the following arguments of Liu et al. (2020) in our notation.

Proof of Lemma 4. The platform uses the true arm rankings π_k. If $U_{j,k}(t) > U_{j,k'}(t)$, then player p_j ranks a_k above $a_{k'}$ under the submitted ranking. The pair (p_j, a_k) would then block m_t under those rankings, contradicting stability. □

Proof of Lemma 2. Fix a blocking triplet $(p_j, a_k, a_{k'})$. Whenever this triplet blocks the selected matching in round t, player p_j receives $a_{k'}$ and observes its reward, and Lemma 4 gives $U_{j,k'}(t) \geq U_{j,k}(t)$. Since $\mu_{j,k'} < \mu_{j,k}$, the standard UCB analysis (Bubeck & Cesa-Bianchi, 2012, Chapter 2) gives

$$\sum_{m \in B_{j,k,k'}} \mathbb{E}T_m(T) \leq 5 + \frac{6 \log T}{\Delta^2_{j,k,k'}}. \qquad \square$$

B Proof of Theorem 5

To prove Theorem 5, fix a player p_j, an assigned arm $a_{k'}$, a threshold $d > 0$, and a horizon T. Let $\mathcal{A}_{j,k'}(d) = \{k : \Delta_{j,k,k'} \geq d,\ B_{j,k,k'} \neq \varnothing\}$, so $|\mathcal{A}_{j,k'}(d)| = H_{j,k'}(d)$. For an event E, let $\mathbf{1}_{\{E\}}$ equal one when E occurs and zero otherwise. The number of rounds to be bounded is

$$L_{j,k'}(d,T) := \sum_{t=1}^{T} \mathbf{1}_{\{m_t \in \mathcal{S}_{j,k'}(d)\}} = \sum_{m \in \mathcal{S}_{j,k'}(d)} T_m(T). \tag{4}$$

If $\mathcal{A}_{j,k'}(d) = \varnothing$, then $L_{j,k'}(d,T) = 0$ and the result is immediate. Hence we assume $\mathcal{A}_{j,k'}(d) \neq \varnothing$.

To control deviations at adaptive observation counts, view each player-arm pair's rewards as an i.i.d. sequence, revealing the next reward whenever the pair is matched. Let $\widehat{\mu}_{j,\ell,s}$ denote the mean of the first s rewards of (p_j, a_ℓ), and define $r_t(s) = \sqrt{3 \log t/(2s)}$ for $t \geq 2$ and $s \geq 1$. Define

$$\begin{aligned} \mathcal{E}^+_{j,\ell}(t) &:= \{\exists\, 1 \leq s \leq t-1 : \widehat{\mu}_{j,\ell,s} - \mu_{j,\ell} > r_t(s)\}, \\ \mathcal{E}^-_{j,\ell}(t) &:= \{\exists\, 1 \leq s \leq t-1 : \widehat{\mu}_{j,\ell,s} - \mu_{j,\ell} < -r_t(s)\}. \end{aligned} \tag{5}$$

Hoeffding's inequality (Hoeffding, 1963, Theorem 1) and a union bound over s give

$$\Pr\big(\mathcal{E}^+_{j,\ell}(t)\big),\ \Pr\big(\mathcal{E}^-_{j,\ell}(t)\big) \leq \sum_{s=1}^{t-1} t^{-3} = (t-1)t^{-3} \leq t^{-2}. \tag{6}$$

Whenever $N_{j,\ell}(t-1) = s > 0$, the empirical mean used by UCB-GS is $\widehat{\mu}_{j,\ell,s}$. Thus a confidence failure at this observation count is contained in the corresponding event in (5).

Set $q = \lfloor 6 \log T/d^2 \rfloor + 1$ and $h = H_{j,k'}(d) + 1$. Here h accounts for one upper deviation event for the assigned arm and one lower deviation event for each arm in $\mathcal{A}_{j,k'}(d)$. The following lemma bounds these rounds using a threshold n on the observation count $N_{j,k'}(t-1)$.

Lemma 15 (Joint count bound at a fixed threshold)**.** *For every integer $n \geq q$, we have $\mathbb{E}L_{j,k'}(d,T) \leq n + h/n$.*

Proof. If $n \geq T$, the deterministic bound $L_{j,k'}(d,T) \leq T \leq n$ proves the claim. Assume $n < T$. Every such round increases $N_{j,k'}$ by one, so at most n such rounds occur while $N_{j,k'}(t-1) < n$.

Now consider such a round with $N_{j,k'}(t-1) \geq n$. Since a pair is observed at most once per round, $N_{j,k'}(t-1) \leq t-1$, so $t \geq n+1$. Also $n \geq q > 6 \log T/d^2$, which implies $2r_t\big(N_{j,k'}(t-1)\big) \leq \sqrt{\frac{6 \log T}{n}} < d$.

Since $m_t \in \mathcal{S}_{j,k'}(d)$, there is some $k \in \mathcal{A}_{j,k'}(d)$ such that (p_j, a_k) blocks m_t. Lemma 4 gives

$$U_{j,k'}(t) \geq U_{j,k}(t). \tag{7}$$

Since $N_{j,k'}(t-1) \geq n \geq 1$, $U_{j,k'}(t)$ is finite. If $N_{j,k}(t-1) = 0$, then $U_{j,k}(t) = +\infty$, contradicting (7). Otherwise, if neither $\mathcal{E}^+_{j,k'}(t)$ nor $\mathcal{E}^-_{j,k}(t)$ occurs, the UCB definition and $\Delta_{j,k,k'} \geq d$ give

$$U_{j,k'}(t) \leq \mu_{j,k'} + 2r_t\big(N_{j,k'}(t-1)\big) < \mu_{j,k'} + d \leq \mu_{j,k} \leq U_{j,k}(t),$$

which is again a contradiction. Consequently, for every $t \geq n+1$,

$$\{m_t \in \mathcal{S}_{j,k'}(d),\ N_{j,k'}(t-1) \geq n\} \subseteq \mathcal{E}^+_{j,k'}(t) \cup \bigcup_{k \in \mathcal{A}_{j,k'}(d)} \mathcal{E}^-_{j,k}(t). \tag{8}$$

By (6) and a union bound, the probability of the right-hand side is at most h/t^2.

Taking expectations in (4) and splitting at the threshold n gives

$$\begin{aligned} \mathbb{E}L_{j,k'}(d,T) &\leq n + \sum_{t=n+1}^{T} \Pr\{m_t \in \mathcal{S}_{j,k'}(d),\ N_{j,k'}(t-1) \geq n\} \\ &\leq n + h \sum_{t=n+1}^{\infty} t^{-2} \leq n + \frac{h}{n}, \end{aligned} \tag{9}$$

where the last inequality uses $\sum_{t=n+1}^{\infty} t^{-2} \leq \int_n^{\infty} x^{-2}\,dx = 1/n$. □

Choose $n = \max\{q, \lceil\sqrt{h}\rceil\}$, so $n \leq q + \sqrt{h} + 1$ and $h/n \leq \sqrt{h}$. Lemma 15 therefore yields

$$\mathbb{E}L_{j,k'}(d,T) \leq n + \frac{h}{n} \leq q + 1 + 2\sqrt{h} = \left\lfloor \frac{6\log T}{d^2} \right\rfloor + 2 + 2\sqrt{H_{j,k'}(d)+1}.$$

Combining this with $L_{j,k'}(d,T) \leq T$ and (4) proves Theorem 5.

C Finite Thresholds for the Joint Constraints

For each assigned pair $(p_j, a_{k'})$, let $\mathcal{D}_{j,k'} = \{\Delta_{j,k,k'} : k \in [K],\ B_{j,k,k'} \neq \varnothing\}$ be the set of positive triplet gaps associated with this pair. The linear program defining $\Psi_i(T)$ in Section 4.2 is

$$\begin{aligned} \Psi_i(T) = \max_{x_m \geq 0} \quad & \sum_{m \in \mathcal{M}} \Delta_{i,m(i)} x_m \\ \text{subject to} \quad & \sum_{m \in B_{j,k,k'}} x_m \leq 5 + \frac{6\log T}{\Delta^2_{j,k,k'}} \quad \text{for every blocking triplet } (p_j, a_k, a_{k'}), \\ & \sum_{m \in \mathcal{S}_{j,k'}(d)} x_m \leq C_{j,k'}(d,T) \quad \text{for all } j \in [N],\ k' \in [K],\ d \in \mathcal{D}_{j,k'}, \\ & \sum_{m \in \mathcal{M}} x_m \leq T. \end{aligned} \tag{10}$$

The following lemma shows that it suffices to impose the joint constraints at these finite thresholds.

Lemma 16 (Finite threshold reduction)**.** *For every $j \in [N]$, $k' \in [K]$, and $d > 0$, either $\mathcal{S}_{j,k'}(d)$ is empty or there is $d^\star \in \mathcal{D}_{j,k'}$ such that $\mathcal{S}_{j,k'}(d) = \mathcal{S}_{j,k'}(d^\star)$ and $C_{j,k'}(d^\star,T) \leq C_{j,k'}(d,T)$. Consequently, imposing the joint constraints at every $d > 0$ gives the same feasible set and optimal value as* (10)*.*

Proof. Fix j, k' and $d > 0$. If $\mathcal{S}_{j,k'}(d)$ is empty, the corresponding constraint holds automatically.

Otherwise, let $d^\star = \min\{d' \in \mathcal{D}_{j,k'} : d' \geq d\}$. For every arm a_k with $B_{j,k,k'} \neq \varnothing$, we have $\Delta_{j,k,k'} \geq d$ if and only if $\Delta_{j,k,k'} \geq d^\star$. The two thresholds therefore select the same blocking arms, giving $\mathcal{S}_{j,k'}(d) = \mathcal{S}_{j,k'}(d^\star)$ and $H_{j,k'}(d) = H_{j,k'}(d^\star)$. Since $d^\star \geq d$, the term $\lfloor 6\log T/(d^\star)^2 \rfloor$ is at most $\lfloor 6\log T/d^2 \rfloor$, so (3) gives $C_{j,k'}(d^\star,T) \leq C_{j,k'}(d,T)$.

For any feasible solution under the finite constraints, the constraint at $d^\star \in \mathcal{D}_{j,k'}$ gives

$$\sum_{m \in \mathcal{S}_{j,k'}(d)} x_m = \sum_{m \in \mathcal{S}_{j,k'}(d^\star)} x_m \le C_{j,k'}(d^\star, T) \le C_{j,k'}(d, T).$$

Conversely, the constraints at every positive threshold include those at $d \in \mathcal{D}_{j,k'}$, so the two formulations have the same feasible set and optimal value. □

D Comparisons between the Regret Bounds

D.1 Comparison with Separate Covers

Proof of Theorem 3. Taking $x_m = \mathbb{E}T_m(T)$ satisfies the triplet constraints by Lemma 2 and the horizon constraint because $\sum_m \mathbb{E}T_m(T) = T$. This feasible solution has objective value $R_i(T)$, so $R_i(T) \le \Gamma_i(T)$.

For each arm a_ℓ with $\Delta_{i,\ell} > 0$, choose a cover $Q_\ell \in \mathcal{C}(\mathcal{M}_{i,\ell})$ minimizing the sum in (1). Any feasible solution to (2) satisfies

$$\sum_{m \in \mathcal{M}_{i,\ell}} x_m \le \sum_{(p_j, a_k, a_{k'}) \in Q_\ell} \sum_{m \in B_{j,k,k'}} x_m \le \sum_{(p_j, a_k, a_{k'}) \in Q_\ell} \left(5 + \frac{6 \log T}{\Delta^2_{j,k,k'}}\right). \tag{11}$$

Every matching with a positive regret gap is unstable. We can therefore bound the objective by discarding its nonpositive terms and grouping the rest over the sets $\mathcal{M}_{i,\ell}$. Multiplying (11) by $\Delta_{i,\ell}$ and summing over arms with $\Delta_{i,\ell} > 0$ gives $\sum_m \Delta_{i,m(i)} x_m \le \Lambda_i(T)$. Maximizing over feasible solutions proves $\Gamma_i(T) \le \Lambda_i(T)$. □

For completeness, we also bound $\Lambda_i(T)$ in terms of $\Gamma_i(T)$.

Proposition 17 (Comparison of Λ_i and Γ_i)**.** *For every fixed market, every player p_i, and all sufficiently large T, $\Lambda_i(T) \le (K-1)\big(1 + (N-1)\log K\big)\Gamma_i(T)$.*

Proof. Fix an arm index ℓ with $\Delta_{i,\ell} > 0$ and $\mathcal{M}_{i,\ell} \neq \varnothing$. Let $\tau = (p_j, a_k, a_{k'})$ range over all blocking triplets with $B_{j,k,k'} \neq \varnothing$, and write $B_\tau = B_{j,k,k'}$ and $b_\tau(T) = 5 + 6\log T/\Delta^2_{j,k,k'}$. A fractional cover assigns a weight $y_\tau \ge 0$ to each triplet, with total weight at least one on the triplets blocking each matching in $\mathcal{M}_{i,\ell}$. Its minimum cost is

$$F_\ell(T) = \min_{y_\tau \ge 0} \left\{ \sum_\tau b_\tau(T) y_\tau : \sum_{\tau : m \in B_\tau} y_\tau \ge 1 \text{ for every } m \in \mathcal{M}_{i,\ell} \right\}.$$

Every matching in $\mathcal{M}_{i,\ell}$ is unstable and hence admits a blocking triplet, so a cover exists. Linear programming duality gives

$$F_\ell(T) = \max_{z_m \ge 0} \left\{ \sum_{m \in \mathcal{M}_{i,\ell}} z_m : \sum_{m \in B_\tau \cap \mathcal{M}_{i,\ell}} z_m \le b_\tau(T) \text{ for every } \tau \right\}.$$

Since every matching in $\mathcal{M}_{i,\ell}$ belongs to at least one blocking triplet set, each z_m appears at least once when the dual constraints are summed. Hence $\sum_{m \in \mathcal{M}_{i,\ell}} z_m \le \sum_\tau b_\tau(T)$. For the fixed market, the right-hand side is $O(\log T)$ and hence at most T for all sufficiently large T. A dual optimizer, extended by zero outside $\mathcal{M}_{i,\ell}$, is therefore feasible for (2), yielding $\Delta_{i,\ell} F_\ell(T) \le \Gamma_i(T)$.

Apply the weighted set cover bound of Chvatal (1979) to the universe $\mathcal{M}_{i,\ell}$, with sets $B_\tau \cap \mathcal{M}_{i,\ell}$ and costs $b_\tau(T)$. Since $F_\ell(T)$ is the corresponding fractional optimum, the minimum integral cover cost is at most $(1 + \log|\mathcal{M}_{i,\ell}|)F_\ell(T)$.

Since the target assignment is fixed, $|\mathcal{M}_{i,\ell}| \le K^{N-1}$. The minimum cover cost is therefore at most $\big(1 + (N-1)\log K\big)F_\ell(T)$. Arms with $\mathcal{M}_{i,\ell} = \varnothing$ contribute zero, and at most $K-1$ arms have positive regret gaps. Multiplying the cover costs by $\Delta_{i,\ell}$, summing as in (1), and using $\Delta_{i,\ell} F_\ell(T) \le \Gamma_i(T)$ proves the result. □

D.2 Sharing Observation Counts

Proof of Proposition 7. The inequality $\Psi_i(T) \le \Gamma_i(T)$ follows from Theorem 6. Fix a feasible solution x to (2) and set $y = x/(2N)$. Scaling preserves nonnegativity, the triplet constraints, and the horizon constraint, so it remains to check the joint constraints.

Fix j, k' and $d > 0$, and write $H = H_{j,k'}(d)$ and $a = 6\log T/d^2$. If $H = 0$, the joint constraint holds automatically, so we assume $H \ge 1$. In each matching in $\mathcal{S}_{j,k'}(d)$, player p_j receives $a_{k'}$, which is not among the H arms counted by $H_{j,k'}(d)$. Hence at most $N-1$ of these H arms are assigned, and every unmatched one forms a blocking pair with p_j. Moreover, membership in $\mathcal{S}_{j,k'}(d)$ guarantees at least one such blocking pair. Therefore, each matching in this union belongs to at least $g = \max\{1, H - N + 1\}$ of the triplet sets defining the union. Summing their constraints gives

$$g \sum_{m \in \mathcal{S}_{j,k'}(d)} x_m \le \sum_{k:\, \Delta_{j,k,k'} \ge d,\ B_{j,k,k'} \ne \varnothing} \ \sum_{m \in B_{j,k,k'}} x_m \le H(5+a).$$

If $H \le N$, then $g = 1$ and $H/g \le N$; otherwise, $H \le N(H - N + 1) = Ng$. Hence

$$\sum_{m \in \mathcal{S}_{j,k'}(d)} y_m \le \frac{5+a}{2} \le \lfloor a \rfloor + 2 + 2\sqrt{H+1},$$

where the last inequality uses $\lfloor a \rfloor \ge a - 1$, $a \ge 0$, and $H \ge 1$. The horizon constraint also gives $\sum_{m \in \mathcal{S}_{j,k'}(d)} y_m \le T/(2N) \le T$. Thus y satisfies the joint constraint with right-hand side $C_{j,k'}(d, T)$ and is feasible for (10). Its objective value is that of x divided by $2N$, so taking x to be an optimizer for $\Gamma_i(T)$ proves $\Psi_i(T) \ge \Gamma_i(T)/(2N)$. □

D.3 Global Preferences

As an additional example, we illustrate the effect of observation count sharing on the market with global preferences in Liu et al. (2020, Example 5). Every player ranks the arms as $a_1 \succ \cdots \succ a_K$, and every arm ranks the players as $p_1 \succ \cdots \succ p_N$. The unique stable matching assigns a_i to p_i, so $\Delta_{i,\ell} = \mu_{i,i} - \mu_{i,\ell}$ is positive exactly when $\ell > i$. The joint constraints give

$$R_i(T) \le \Psi_i(T) \le \left(2 + 2\sqrt{i+1}\right) \sum_{\ell=i+1}^{K} \Delta_{i,\ell} + 6\log T \sum_{\ell=i+1}^{K} \frac{1}{\Delta_{i,\ell}}. \tag{12}$$

Compared with equation (8) of Liu et al. (2020), this bound removes the factor i from the logarithmic term.

Whenever p_i receives a_ℓ with $\ell > i$, at least one of $a_1, \dots, a_i$ is unmatched or held by a player ranked below p_i, since only $i-1$ players rank above p_i. That arm forms a blocking pair with p_i and has gap at least $\Delta_{i,\ell}$. Thus $\mathcal{M}_{i,\ell} \subseteq \mathcal{S}_{i,\ell}(\Delta_{i,\ell})$. Only arms among $a_1, \dots, a_i$ can meet this gap threshold, so $H_{i,\ell}(\Delta_{i,\ell}) \le i$. By Lemma 16, every feasible solution to (10) satisfies the joint constraint at $d = \Delta_{i,\ell}$, giving

$$\sum_{m \in \mathcal{M}_{i,\ell}} x_m \le \frac{6\log T}{\Delta_{i,\ell}^2} + 2 + 2\sqrt{i+1}.$$

Discarding nonpositive objective terms, multiplying by $\Delta_{i,\ell}$, and summing over $\ell > i$ bounds the objective of every feasible solution by the right-hand side of (12). Maximizing gives the bound on $\Psi_i(T)$, and Theorem 6 gives $R_i(T) \le \Psi_i(T)$.

The following proposition gives markets with global preferences on which, for sufficiently large T, triplet count sharing leaves the bound unchanged, while observation count sharing improves it by a factor of $\Theta(i)$.

Proposition 18. *For every integer $i \ge 2$ and every $d \in (0, 1/4]$, there is a market with $N = K = i+1$ and global preferences such that $\Delta_{i,i+1} = d$ and, for all sufficiently large T,*

$$\Lambda_i(T) = \Gamma_i(T) = \Theta\left(\frac{i\log T}{d}\right), \qquad \Psi_i(T) = \Theta\left(\frac{\log T}{d}\right).$$

Consequently, $\Gamma_i(T)/\Psi_i(T) = \Theta(i)$. The implicit constants are independent of i and d.

Proof. Set $\varepsilon = d/i$ and choose the mean rewards

$$\mu_{j,k} = \begin{cases} \frac{1}{2} + d + (j-k)\varepsilon, & k \le j, \\ \frac{1}{2} - (k-j-1)\varepsilon, & k > j, \end{cases} \qquad j, k \in [i+1].$$

These means lie in $[0,1]$ and strictly decrease with the arm index. Let every arm rank $p_1 \succ \cdots \succ p_{i+1}$. The unique stable matching assigns a_j to p_j. For the target p_i, only a_{i+1} has a positive regret gap, equal to d. Here the minimum reward gap in Definition 1 is $\Delta = \varepsilon$.

For each $r \in [i]$, define a matching $m^{(r)}$ by

$$m^{(r)}(j) = \begin{cases} j, & j < r, \\ j+1, & r \le j \le i, \\ r, & j = i+1. \end{cases}$$

Each matching assigns a_{i+1} to p_i. For every $r \le j \le i$, the pair (p_j, a_r) blocks $m^{(r)}$, since a_r is held by p_{i+1}. Every other arm that p_j prefers to a_{j+1} is held by a player ranked above p_j. For every player outside this range, every arm preferred to its assignment is held by a player ranked above it. The complete set of blocking triplets of $m^{(r)}$ is therefore

$$\big\{(p_j, a_r, a_{j+1}) : r \le j \le i\big\}.$$

These sets are pairwise disjoint as r varies.

The gap of (p_j, a_r, a_{j+1}) is $\Delta_{j,r,j+1} = d + (j-r)\varepsilon$. For each r, the target's triplet (p_i, a_r, a_{i+1}) has the largest gap among the triplets blocking $m^{(r)}$ and hence the smallest cost. Write $g_r = d + (i-r)\varepsilon$ and $b_r(T) = 5 + 6\log T/g_r^2$. Thus every triplet that blocks $m^{(r)}$ has cost at least $b_r(T)$. Every cover of $\mathcal{M}_{i,i+1}$ must contain a triplet blocking each $m^{(r)}$. Since their blocking triplet sets are disjoint, every such cover costs at least $\sum_{r=1}^{i} b_r(T)$. Conversely, the i target triplets (p_i, a_r, a_{i+1}), $r \in [i]$, cover $\mathcal{M}_{i,i+1}$ by the argument preceding the proposition. Their total cost attains this lower bound, so $\Lambda_i(T) = d\sum_{r=1}^{i} b_r(T)$.

To obtain the same lower bound for $\Gamma_i(T)$, set $x_{m^{(r)}} = b_r(T)$ for every $r \in [i]$ and all other variables to zero. Each triplet constraint contains at most one of these nonzero counts. If it contains $x_{m^{(r)}}$, its gap is at most g_r, so its right-hand side is at least $b_r(T)$. Thus all triplet constraints hold. For fixed i and d, the total count $\sum_{r=1}^{i} b_r(T)$ is $O(\log T)$ and hence at most T for all sufficiently large T. The resulting solution is feasible for (2) and has objective value $\Lambda_i(T)$. Since $\Gamma_i(T) \le \Lambda_i(T)$, we obtain

$$\Gamma_i(T) = \Lambda_i(T) = d\sum_{r=1}^{i}\left(5 + \frac{6\log T}{[d + (i-r)\varepsilon]^2}\right).$$

The inequalities $d \le g_r < 2d$ give $\Gamma_i(T) = \Theta(i\log T/d)$.

For $\Psi_i(T)$, equation (12) gives $\Psi_i(T) \le \big(2 + 2\sqrt{i+1}\big)d + \frac{6\log T}{d}$. For the lower bound on $\Psi_i(T)$, note that $m^{(i)}$ has only one blocking triplet, (p_i, a_i, a_{i+1}), whose gap is d. For sufficiently large T, set $x_{m^{(i)}} = 6\log T/d^2 \le T$ and all other variables to zero. The triplet and horizon constraints hold. Any joint constraint containing $m^{(i)}$ has assigned pair (p_i, a_{i+1}) and a threshold $u \le d$. Since $\lfloor z \rfloor + 2 \ge z$, its right-hand side satisfies

$$C_{i,i+1}(u,T) \ge \min\left\{T, \frac{6\log T}{u^2}\right\} \ge \frac{6\log T}{d^2}.$$

Thus every joint constraint is satisfied. This feasible solution has objective value $6\log T/d$, yielding

$$\frac{6\log T}{d} \le \Psi_i(T) \le \frac{6\log T}{d} + \big(2 + 2\sqrt{i+1}\big)d.$$

For all sufficiently large T, the additive term is at most $\log T/d$. Hence $\Psi_i(T) = \Theta(\log T/d)$, and combining the two bounds gives $\Gamma_i(T)/\Psi_i(T) = \Theta(i)$. □

E BLOCKING PAIRS INVOLVING THE FIRST N CHOICES

We establish two rank properties and construct a cover of unstable matchings using blocking arms among each player's first N choices. We then bound the sum of triplet count bounds for each assigned pair.

Lemma 19. *Every stable arm of a player is among that player's first N choices.*

Proof. Suppose a stable matching assigns a player an arm outside its first N choices. At most $N-1$ of the player's first N choices are assigned to other players, so at least one is unmatched and forms a blocking pair with this player, contradicting stability. □

Lemma 20. *Every unstable matching has a blocking pair (p_j, a_k) for which a_k is among player p_j's first N choices.*

Proof. If some player p_j receives an arm outside its first N choices, at most $N-1$ of those choices are assigned to other players. An unmatched one therefore forms a blocking pair with p_j. Otherwise, every player receives one of its first N choices. Since the matching is unstable, it has a blocking pair (p_j, a_k). Its blocking arm is preferred to p_j's assignment and is therefore also among p_j's first N choices. □

For a blocking triplet $(p_j, a_k, a_{k'})$ whose blocking arm is among p_j's first N choices, let $r \le N$ and $s > r$ be the ranks of a_k and $a_{k'}$, respectively. Then $\Delta_{j,k,k'} = \sum_{q=r}^{s-1} \left(\mu_{j,\sigma_{j,q}} - \mu_{j,\sigma_{j,q+1}}\right) \ge \Delta$. Indeed, $r < s \le K$ implies $r \le \min\{N, K-1\}$, so the first term is at least Δ and all remaining terms are positive.

Let $\mathrm{Top}_N(j) = \{\sigma_{j,1}, \ldots, \sigma_{j,N}\}$ be the indices of player p_j's first N choices, and define

$$\mathcal{A}^N_{j,k'} = \{k \in \mathrm{Top}_N(j) : B_{j,k,k'} \neq \varnothing\}, \qquad \mathcal{I}_N = \left\{(p_j, a_{k'}) : \mathcal{A}^N_{j,k'} \neq \varnothing\right\}.$$

For $(p_j, a_{k'}) \in \mathcal{I}_N$, set $d^N_{j,k'} = \min_{k \in \mathcal{A}^N_{j,k'}} \Delta_{j,k,k'}$ and $H^N_{j,k'} = |\mathcal{A}^N_{j,k'}|$.

Lemma 21 (Threshold cover from the first N choices)**.** *For every $(p_j, a_{k'}) \in \mathcal{I}_N$, we have $d^N_{j,k'} \in \mathcal{D}_{j,k'}$ and $d^N_{j,k'} \ge \Delta$. Moreover, $\mathcal{S}_{j,k'}(d^N_{j,k'}) = \bigcup_{k \in \mathcal{A}^N_{j,k'}} B_{j,k,k'}$ and $H_{j,k'}(d^N_{j,k'}) = H^N_{j,k'} \le N$. The index set satisfies $|\mathcal{I}_N| \le N(K-1)$, and $\{m : m \text{ is unstable}\} \subseteq \bigcup_{(p_j, a_{k'}) \in \mathcal{I}_N} \mathcal{S}_{j,k'}(d^N_{j,k'})$.*

Proof. Fix $(p_j, a_{k'}) \in \mathcal{I}_N$ and choose $k_0 \in \mathcal{A}^N_{j,k'}$ attaining $d^N_{j,k'}$. Then $d^N_{j,k'} = \Delta_{j,k_0,k'} \in \mathcal{D}_{j,k'}$ because $B_{j,k_0,k'} \neq \varnothing$, and $d^N_{j,k'} \ge \Delta$ by the bound above. Every $k \in \mathcal{A}^N_{j,k'}$ has gap at least $d^N_{j,k'}$ and is therefore counted by $H_{j,k'}(d^N_{j,k'})$. Conversely, any k counted by $H_{j,k'}(d^N_{j,k'})$ satisfies $B_{j,k,k'} \neq \varnothing$ and $\mu_{j,k} - \mu_{j,k'} \ge d^N_{j,k'} = \mu_{j,k_0} - \mu_{j,k'}$, so $\mu_{j,k} \ge \mu_{j,k_0}$. Thus a_k ranks at least as highly as a_{k_0} and is also among p_j's first N choices, giving $k \in \mathcal{A}^N_{j,k'}$. The two index sets therefore coincide, giving the stated identities for $\mathcal{S}_{j,k'}(d^N_{j,k'})$ and $H_{j,k'}(d^N_{j,k'})$. Also, $H^N_{j,k'} \le N$ since $\mathrm{Top}_N(j)$ has N elements. A player's most preferred arm cannot be the assigned arm of a blocking triplet. Each player therefore appears in at most $K-1$ pairs in $\mathcal{I}_N$, giving $|\mathcal{I}_N| \le N(K-1)$. Finally, for any unstable matching m, Lemma 20 gives a blocking triplet $(p_j, a_k, a_{k'})$ with $k \in \mathrm{Top}_N(j)$ and $k' = m(j)$. Since $m \in B_{j,k,k'}$, we have $k \in \mathcal{A}^N_{j,k'}$ and $(p_j, a_{k'}) \in \mathcal{I}_N$. Therefore, $m \in \mathcal{S}_{j,k'}(d^N_{j,k'})$. □

To bound $\Gamma_i(T)$ using triplet constraints, we also need to sum the triplet count bounds for each assigned pair.

Lemma 22 (Triplet count bounds for one assigned pair)**.** *Let $D = \min\{N, K-1\}$. For every assigned pair $(p_j, a_{k'})$,*

$$\sum_{k \in \mathcal{A}^N_{j,k'}} \left(5 + \frac{6 \log T}{\Delta^2_{j,k,k'}}\right) \le 5D + \frac{\pi^2 \log T}{\Delta^2} = O\left(\frac{\log T}{\Delta^2}\right).$$

Proof. Let v be the rank of $a_{k'}$ and set $u = \min\{v-1, N\} \le D$. Each arm indexed by $\mathcal{A}^N_{j,k'}$ has rank $r \le u$. Since $u+1 \le v$,

$$\Delta_{j,\sigma_{j,r},k'} \ge \sum_{q=r}^{u} (\mu_{j,\sigma_{j,q}} - \mu_{j,\sigma_{j,q+1}}) \ge (u-r+1)\Delta,$$

where the last inequality follows from Definition 1 since $u \le D$. The integers $u-r+1$ are distinct as the arm varies, and therefore

$$\sum_{k \in \mathcal{A}^N_{j,k'}} \frac{1}{\Delta^2_{j,k,k'}} \le \frac{1}{\Delta^2} \sum_{h=1}^{u} \frac{1}{h^2} \le \frac{\pi^2}{6\Delta^2}.$$

Multiplying by $6 \log T$ and using $|\mathcal{A}^N_{j,k'}| \le u \le D$ proves the first inequality. Finally, the first D adjacent gaps sum to at most one, so $D\Delta \le 1$ and hence $5D \le 5/\Delta = O(\log T/\Delta^2)$. □

F Proof of the General Market Regret Bound

We first prove Proposition 9 and use its three classes to prove Theorem 8. We then derive a finite-time bound for $\Psi_i(T)$ using the joint constraints and a finer bound using a separate minimum reward gap for each player.

Let $S = \{a_{\underline{m}(j)} : j \in [N]\}$. We call $a_{\underline{m}(j)}$ player p_j's reference arm and call arms outside S external arms, which are unmatched under $\underline{m}$. Fix a target player p_i, and let $\mathcal{M}_i^{(r)}$ denote the r-th class in Proposition 9. If the third class is nonempty, we associate each matching in this class with a stable matching in the market restricted to S. Taking the player join of these restricted stable matchings gives a stable matching that still assigns the target an arm worse than its player-pessimal stable arm. Its extension to the full market must therefore admit a blocking pair with an arm outside S, which identifies an external arm assigned in every matching of the third class.

F.1 A Stable Matching on the Reference Arms

Restrict the market to all players and the arms in S, keeping the original preferences over these agents. The matching $\underline{m}$ remains stable in this restricted market, since removing unmatched arms cannot create a blocking pair. Deleting the assignments to external arms from a matching in the third class leaves a partial matching on S. Any player who remains matched keeps its assignment, so a blocking pair involving that player would also block the original matching, contradicting the definition of the third class. Thus only unmatched players can belong to blocking pairs in the partial matching. We combine it with $\underline{m}$ by letting each arm in S choose its preferred holder. This is a special case of the construction in Blum et al. (1997, Theorem A.3), with players corresponding to firms and arms to workers. We give the direct argument below.

Lemma 23 (Restricted stable matching)**.** *For every matching m in the third class, there is a matching ν_m that is stable in the restricted market and satisfies $\nu_m(i) = m(i)$. Each player receives either its arm under m or its player-pessimal stable arm. In particular, $a_{m(i)} \in S \setminus \{a_{\underline{m}(i)}\}$.*

Proof. For each arm in S, choose its preferred holder from $\underline{m}$ and m. If the arm is unmatched in m, choose its holder under $\underline{m}$. These choices define a matching if no player is chosen twice.

Suppose two distinct arms choose the same player p_j. The only arms that can choose p_j are $a_{\underline{m}(j)}$ and, if $a_{m(j)} \in S$, $a_{m(j)}$. Thus these are the two distinct arms, and $a_{m(j)} \in S$. If p_j prefers $a_{\underline{m}(j)}$ to $a_{m(j)}$, the choice made by $a_{\underline{m}(j)}$ means that it is unmatched in m or prefers p_j to its holder under m. Together with the player's preference, this makes $(p_j, a_{\underline{m}(j)})$ a blocking pair for m. Since p_j holds an arm in S, this contradicts membership in the third class. If instead p_j prefers $a_{m(j)}$ to $a_{\underline{m}(j)}$, the choice made by $a_{m(j)}$ means that it prefers p_j to its holder under $\underline{m}$. That holder is different from p_j, because $a_{m(j)} \ne a_{\underline{m}(j)}$. The pair $(p_j, a_{m(j)})$ would then block $\underline{m}$, contradicting stability. Thus no player is chosen twice. The N arms choose N distinct players, giving a matching ν_m that assigns an arm to every player. By construction, $\nu_m(j) \in \{m(j), \underline{m}(j)\}$ for every j. Every arm weakly prefers its new holder to its holder under $\underline{m}$ and, whenever it is matched in m, to its holder under m.

Suppose (p_j, a) blocks ν_m, where $a \in S$. If $\nu_m(j) = \underline{m}(j)$, player p_j prefers a to its player-pessimal stable arm. Arm a prefers p_j to its new holder, which it weakly prefers to its holder under $\underline{m}$. Hence the same pair blocks $\underline{m}$, a contradiction. Otherwise $\nu_m(j) = m(j)$, so $a_{m(j)} \in S$ and p_j prefers a to its original assignment. If a is matched in m, it prefers p_j to its holder under m because its new holder is weakly preferred to that holder. If a is unmatched in m, it accepts p_j. In either case, the pair blocks m while p_j holds an arm in S, a contradiction. Hence ν_m is stable in the restricted market.

Finally, p_i prefers $a_{\underline{m}(i)}$ to its current arm, but this pair does not block m. The arm $a_{\underline{m}(i)}$ must therefore have a holder under m whom it prefers to p_i, and the construction does not assign it to p_i. Since $\nu_m(i) \in \{m(i), \underline{m}(i)\}$ and the construction rules out $\nu_m(i) = \underline{m}(i)$, we have $\nu_m(i) = m(i)$. Because ν_m is a matching on S, this also implies $a_{m(i)} \in S$. The positive regret gap excludes $a_{\underline{m}(i)}$, giving $a_{m(i)} \in S \setminus \{a_{\underline{m}(i)}\}$. □

Lemma 24. *Let m belong to the third class. If p_q is assigned to an arm outside S, then $(p_q, a_{\underline{m}(q)})$ blocks m.*

Proof. In the construction of Lemma 23, the only arm in S that can choose p_q is $a_{\underline{m}(q)}$. The arm assigned to p_q under m lies outside S and therefore makes no choice in the construction. Since the constructed matching assigns an arm to every player, $a_{\underline{m}(q)}$ must choose p_q. The arm $a_{\underline{m}(q)}$ cannot already be assigned to p_q in m, so it is unmatched in m or strictly prefers p_q to its holder under m.

The current arm $a_{m(q)}$ is unmatched under $\underline{m}$ and would accept p_q. If p_q preferred it to $a_{\underline{m}(q)}$, the pair $(p_q, a_{m(q)})$ would block $\underline{m}$. Stability and strict player preferences therefore give $a_{\underline{m}(q)} \succ_q a_{m(q)}$. Both conditions for $(p_q, a_{\underline{m}(q)})$ to block m hold. □

F.2 A Common External Arm

Assume that $\mathcal{M}_i^{(3)}$ is nonempty. Lemma 23 gives a restricted stable matching ν_m for every $m \in \mathcal{M}_i^{(3)}$, with $\nu_m(i) = m(i)$.

We write $a \succeq_j b$ when $a = b$ or $a \succ_j b$. For two stable matchings in this restricted market, the player join assigns each player its preferred arm from the two matchings and is itself stable (Roth & Sotomayor, 1990; Blum et al., 1997, Theorem 2.1). Repeatedly taking this join over the finitely many matchings ν_m gives a restricted stable matching $\nu^\star$ satisfying

$$a_{\nu^\star(j)} \succeq_j a_{\nu_m(j)} \quad \text{for every } j \in [N] \text{ and every } m \in \mathcal{M}_i^{(3)}.$$

The target receives its most preferred arm among $\{a_{m(i)} : m \in \mathcal{M}_i^{(3)}\}$. Each of these arms is strictly worse than $a_{\underline{m}(i)}$, so $a_{\underline{m}(i)} \succ_i a_{\nu^\star(i)}$.

Lemma 25 (Common external arm)**.** *If $\mathcal{M}_i^{(3)}$ is nonempty, then some arm $a_{b_i} \notin S$ is assigned in every matching in $\mathcal{M}_i^{(3)}$. The arm also satisfies*

$$\mathcal{M}_i^{(3)} \subseteq \bigcup_{q \in [N] \setminus \{i\}} B_{q, \underline{m}(q), b_i}. \tag{13}$$

Proof. Extend $\nu^\star$ to the full market by keeping every player's assignment in S and leaving all arms outside S unmatched. The extension gives p_i an arm worse than its player-pessimal stable arm. If the extension were stable, this would contradict the definition of the player-pessimal stable arm. Hence the extension is unstable and admits a blocking pair. No pair using an arm in S can block, because $\nu^\star$ is stable in the restricted market and the extension changes none of its assignments. Fix such a blocking pair (p_r, a_{b_i}) with $a_{b_i} \notin S$.

For every $m \in \mathcal{M}_i^{(3)}$, the join property gives $a_{b_i} \succ_r a_{\nu^\star(r)} \succeq_r a_{\nu_m(r)}$.

Fix any such matching m. Lemma 23 gives $a_{\nu_m(r)} \in \{a_{m(r)}, a_{\underline{m}(r)}\}$. If $a_{\nu_m(r)} = a_{\underline{m}(r)}$, then $a_{b_i} \succ_r a_{\underline{m}(r)}$, so p_r prefers a_{b_i} to its player-pessimal stable arm. Since a_{b_i} is unmatched under $\underline{m}$ and accepts p_r, the pair would block $\underline{m}$, a contradiction. Thus $\nu_m(r) = m(r)$ and $a_{m(r)} \in S$. In particular, p_r prefers a_{b_i} to its assignment in the original matching m.

If a_{b_i} were unmatched in m, then (p_r, a_{b_i}) would block m while p_r holds an arm in S. This would contradict membership in the third class. Thus a_{b_i} is assigned in every such m.

Let p_q be the holder of a_{b_i} in m. Since p_q is assigned to an arm outside S, Lemma 24 shows that $(p_q, a_{\underline{m}(q)})$ blocks m. Together with $m(q) = b_i$, this gives $m \in B_{q,\underline{m}(q),b_i}$. Lemma 23 gives $a_{m(i)} \in S$, so $q \neq i$. The arm a_{b_i} is fixed across the third class, but its holder p_q may vary with m. Taking the union over all $q \in [N] \setminus \{i\}$ gives (13). □

Proof of Proposition 9. Given $m \in \mathcal{M}_i^{(2)}$, let p_j be any player who holds an arm $a_{k'} \in S$ and forms a blocking pair. If $a_{k'}$ is among p_j's first N choices, any arm forming a blocking pair with p_j is preferred to $a_{k'}$ and is therefore also among these choices. Otherwise, at most $N-1$ of these choices are assigned to other players, so an unmatched one forms a blocking pair with p_j. If the third class is nonempty, Lemma 25 gives a common external arm assigned in every matching of the class, and Lemma 24 shows that its holder forms a blocking pair with its player-pessimal stable arm. □

F.3 Proof of the General Bound Using Triplet Constraints

Proof of Theorem 8. Let $D = \min\{N, K-1\}$ and fix any feasible solution $(x_m)_{m\in\mathcal{M}}$ to (2). Since $x_m \geq 0$, we discard nonpositive objective terms and bound the contributions of the three classes in Proposition 9.

In the first class, fix an arm $a_{k'}$ assigned to the target in this class. These matchings are blocked by the triplet $(p_i, a_{\underline{m}(i)}, a_{k'})$, whose gap equals the regret gap $\Delta_{i,k'} \in [\Delta, 1]$. The triplet constraint gives

$$\sum_{\substack{m\in\mathcal{M}_i^{(1)}\\ m(i)=k'}} \Delta_{i,m(i)} x_m \leq \Delta_{i,k'} \sum_{m\in B_{i,\underline{m}(i),k'}} x_m \leq 5\Delta_{i,k'} + \frac{6\log T}{\Delta_{i,k'}} \leq 5 + \frac{6\log T}{\Delta}.$$

Since $k' \neq \underline{m}(i)$, summing over at most $K-1$ assigned arms bounds the first class's contribution by $(K-1)(5 + 6\log T/\Delta)$.

For the second class, let $\mathcal{I}_{N,S} = \{(p_j, a_{k'}) \in \mathcal{I}_N : a_{k'} \in S\}$, which has at most N^2 elements. By Proposition 9, every $m \in \mathcal{M}_i^{(2)}$ belongs to some $B_{j,k,k'}$ with $a_{k'} \in S$ and $k \in \mathrm{Top}_N(j)$, so $k \in \mathcal{A}_{j,k'}^N$ and $(p_j, a_{k'}) \in \mathcal{I}_{N,S}$. Since each regret gap is at most one, summing the triplet constraints and applying Lemma 22 gives

$$\sum_{m\in\mathcal{M}_i^{(2)}} \Delta_{i,m(i)} x_m \leq \sum_{(p_j,a_{k'})\in\mathcal{I}_{N,S}} \sum_{k\in\mathcal{A}_{j,k'}^N} \sum_{m\in B_{j,k,k'}} x_m \leq N^2\left(5D + \frac{\pi^2\log T}{\Delta^2}\right).$$

The third class contributes zero if it is empty. Otherwise, Lemma 25 gives the common external arm a_{b_i} and the cover (13). For each $q \neq i$, the arm a_{b_i} is unmatched under $\underline{m}$, so stability gives $a_{\underline{m}(q)} \succ_q a_{b_i}$. Since $a_{\underline{m}(q)}$ is among p_q's first N choices by Lemma 19, Definition 1 gives $\Delta_{q,\underline{m}(q),b_i} \geq \Delta$. The triplet constraints then yield

$$\sum_{m\in\mathcal{M}_i^{(3)}} \Delta_{i,m(i)} x_m \leq \sum_{q\in[N]\setminus\{i\}} \sum_{m\in B_{q,\underline{m}(q),b_i}} x_m \leq (N-1)\left(5 + \frac{6\log T}{\Delta^2}\right).$$

Adding the three contributions bounds the objective. The horizon constraint also gives $\sum_m \Delta_{i,m(i)} x_m \leq \sum_m x_m \leq T$. Maximizing over feasible solutions gives

$$\begin{aligned}\Gamma_i(T) \leq \min\Bigg\{ T,\, (K-1)\left(5 + \frac{6\log T}{\Delta}\right) + N^2\left(5D + \frac{\pi^2\log T}{\Delta^2}\right) \\ + (N-1)\left(5 + \frac{6\log T}{\Delta^2}\right)\Bigg\}.\end{aligned} \tag{14}$$

Lemma 22 also gives $5D + \pi^2\log T/\Delta^2 = O(\log T/\Delta^2)$. The remaining constant terms are absorbed into $K\log T/\Delta$ and $N^2\log T/\Delta^2$, yielding $\Gamma_i(T) = O(K\log T/\Delta + N^2\log T/\Delta^2)$. Theorem 3 then gives the claimed regret bound. □

F.4 A FINITE-TIME BOUND FOR $\Psi_i(T)$

Theorem 26. *For every target player p_i,*

$$R_i(T) \le \Psi_i(T) \le \min\Bigg\{ T,\, N(K-1)\left(\left\lfloor \frac{6\log T}{\Delta^2}\right\rfloor + 2 + 2\sqrt{N+1}\right), \\ (K-1)\left(5+\frac{6\log T}{\Delta}\right) + N^2\left(\left\lfloor \frac{6\log T}{\Delta^2}\right\rfloor + 2 + 2\sqrt{N+1}\right) \\ + (N-1)\left(5+\frac{6\log T}{\Delta^2}\right)\Bigg\}. \tag{15}$$

Proof. Fix any feasible solution $(x_m)_{m\in\mathcal{M}}$ to (10). It satisfies all the triplet constraints, so the first and third classes contribute at most $(K-1)(5+6\log T/\Delta)$ and $(N-1)(5+6\log T/\Delta^2)$, respectively, by the preceding proof.

For the second class, use the same cover indexed by $\mathcal{I}_{N,S}$ and apply the joint constraints. There are at most N^2 assigned pairs, and Lemma 21 shows that each union in this cover equals $\mathcal{S}_{j,k'}(d^N_{j,k'})$ with $d^N_{j,k'} \in \mathcal{D}_{j,k'}$, $d^N_{j,k'} \ge \Delta$, and $H_{j,k'}(d^N_{j,k'}) \le N$. Since $x_m \ge 0$ and each regret gap is at most one,

$$\begin{aligned}\sum_{m\in\mathcal{M}_i^{(2)}} \Delta_{i,m(i)} x_m &\le \sum_{(p_j,a_{k'})\in\mathcal{I}_{N,S}} \sum_{m\in\mathcal{S}_{j,k'}(d^N_{j,k'})} x_m \\ &\le \sum_{(p_j,a_{k'})\in\mathcal{I}_{N,S}} C_{j,k'}(d^N_{j,k'},T) \\ &\le N^2\left(\left\lfloor \frac{6\log T}{\Delta^2}\right\rfloor + 2 + 2\sqrt{N+1}\right).\end{aligned}$$

Discarding nonpositive objective terms and adding the three contributions gives the corresponding bound in (15).

Alternatively, every matching with a positive regret gap is unstable, so we can apply the cover in Lemma 21 without dividing into classes. Since $d^N_{j,k'} \in \mathcal{D}_{j,k'}$ for every pair in $\mathcal{I}_N$, the joint constraints give

$$\begin{aligned}\sum_{m} \Delta_{i,m(i)} x_m &\le \sum_{m:\,\Delta_{i,m(i)}>0} x_m \\ &\le \sum_{(p_j,a_{k'})\in\mathcal{I}_N} \sum_{m\in\mathcal{S}_{j,k'}(d^N_{j,k'})} x_m \\ &\le N(K-1)\left(\left\lfloor \frac{6\log T}{\Delta^2}\right\rfloor + 2 + 2\sqrt{N+1}\right).\end{aligned}$$

Finally, the objective is at most $\sum_m x_m \le T$ by the reward range and the horizon constraint. Maximizing over feasible solutions gives the stated bound on $\Psi_i(T)$, and Theorem 6 gives $R_i(T) \le \Psi_i(T)$. □

The joint constraints combine triplet counts within each assigned pair. For the second class of Proposition 9, our analysis still sums a bound of $O(\log T/\Delta^2)$ over at most N^2 assigned pairs, so the resulting contribution retains the quadratic dependence on N. Exploiting further structure across these assigned pairs may improve the dependence on N.

F.5 PLAYER-SPECIFIC GAPS

Define $\Delta_j = \min_{1\le r\le D}\left(\mu_{j,\sigma_{j,r}} - \mu_{j,\sigma_{j,r+1}}\right)$ for each player p_j, where $D = \min\{N, K-1\}$. Thus $\Delta = \min_{j\in[N]} \Delta_j$. Keeping these gaps separate in the three classes gives the following refinement.

Corollary 27. *For every target player p_i,*

$$R_i(T) \le \Gamma_i(T) \le O\left(\sum_{\ell:\Delta_{i,\ell}>0} \frac{\log T}{\Delta_{i,\ell}} + N\sum_{j=1}^{N} \frac{\log T}{\Delta_j^2}\right).$$

Proof. Fix any feasible solution $(x_m)_{m\in\mathcal{M}}$ to (2). Discard nonpositive objective terms and use the three classes of Proposition 9.

In the first class, every matching that assigns a_ℓ to p_i belongs to $B_{i,\underline{m}(i),\ell}$, whose triplet gap equals $\Delta_{i,\ell}$. Since $a_{\underline{m}(i)}$ is among p_i's first N choices by Lemma 19, summing the adjacent gaps gives $\Delta_{i,\ell} \ge \Delta_i$ whenever $\Delta_{i,\ell} > 0$. Keeping the individual regret gaps in the triplet constraints gives

$$\sum_{m\in\mathcal{M}_i^{(1)}} \Delta_{i,m(i)} x_m \le \sum_{\ell:\Delta_{i,\ell}>0} \Delta_{i,\ell} \sum_{m\in B_{i,\underline{m}(i),\ell}} x_m \le \sum_{\ell:\Delta_{i,\ell}>0} \left(5\Delta_{i,\ell} + \frac{6\log T}{\Delta_{i,\ell}}\right).$$

For the second class, fix an assigned pair $(p_j, a_{k'})$. The proof of Lemma 22 uses only the adjacent gaps of p_j, which are at least Δ_j over the relevant ranks. Applying that argument with Δ_j gives

$$\sum_{k\in\mathcal{A}_{j,k'}^N} \left(5 + \frac{6\log T}{\Delta_{j,k,k'}^2}\right) \le 5D + \frac{\pi^2 \log T}{\Delta_j^2}.$$

By Proposition 9, every matching in the second class has a blocking triplet with its assigned arm in S and its blocking arm among the same player's first N choices. For each player there are at most $|S| = N$ such assigned arms. Using regret gaps at most one and summing the triplet constraints therefore gives

$$\sum_{m\in\mathcal{M}_i^{(2)}} \Delta_{i,m(i)} x_m \le \sum_{j=1}^{N} \sum_{k':a_{k'}\in S} \sum_{k\in\mathcal{A}_{j,k'}^N} \sum_{m\in B_{j,k,k'}} x_m \le N\sum_{j=1}^{N} \left(5D + \frac{\pi^2\log T}{\Delta_j^2}\right).$$

The third class contributes zero if it is empty. Otherwise, Lemma 25 gives a common external arm a_{b_i} and the cover (13). For each $q \ne i$, stability of $\underline{m}$ and the fact that a_{b_i} is unmatched under $\underline{m}$ imply $a_{\underline{m}(q)} \succ_q a_{b_i}$. The reference arm $a_{\underline{m}(q)}$ is among p_q's first N choices, so summing the adjacent gaps and using Δ_q gives $\Delta_{q,\underline{m}(q),b_i} \ge \Delta_q$. The triplet constraints yield

$$\sum_{m\in\mathcal{M}_i^{(3)}} \Delta_{i,m(i)} x_m \le \sum_{q\in[N]\setminus\{i\}} \sum_{m\in B_{q,\underline{m}(q),b_i}} x_m \le \sum_{q\in[N]\setminus\{i\}} \left(5 + \frac{6\log T}{\Delta_q^2}\right).$$

For each player p_j, the first D adjacent gaps sum to at most one, so $D\Delta_j \le 1$. Therefore, each term $5D + \pi^2\log T/\Delta_j^2$ is $O(\log T/\Delta_j^2)$. The third class contributes $O(\sum_{q\ne i}\log T/\Delta_q^2)$, which is absorbed into $N\sum_j \log T/\Delta_j^2$. For every positive regret gap, $\Delta_{i,\ell} \le 1$ also gives $5\Delta_{i,\ell} + 6\log T/\Delta_{i,\ell} = O(\log T/\Delta_{i,\ell})$. Adding the three contributions and maximizing over feasible solutions proves the stated bound on $\Gamma_i(T)$. Theorem 3 gives $R_i(T) \le \Gamma_i(T)$. □

Since every positive target regret gap is at least Δ_i, the first term is at most $(K-1)\log T/\Delta_i$. Using $\Delta_j \ge \Delta$ for every player then recovers the bound in Theorem 8.

G Sharing Triplet Counts

This appendix proves Theorem 11.

G.1 Market Construction

The market M_K^{A} has two players and K arms, where $K \geq 4$. Every arm ranks $p_1 \succ p_2$, and rewards are independent Bernoulli variables with means

$$\begin{aligned} \mu_{1,1} &= \frac{1}{2}, \quad \mu_{1,2} = \frac{1}{2} - \frac{1}{K}, \quad \mu_{1,\ell} = \frac{1}{K\ell} \quad (\ell \geq 3), \\ \mu_{2,1} &= \frac{1}{K}, \quad \mu_{2,2} = \frac{1}{2}, \qquad\quad \mu_{2,\ell} = \frac{1}{K\ell} \quad (\ell \geq 3). \end{aligned} \tag{16}$$

For $K \geq 4$, all means lie in $(0, 1)$ and are distinct for each player. Player p_1 ranks $a_1 \succ a_2 \succ a_3 \succ \cdots \succ a_K$, and player p_2 ranks $a_2 \succ a_1 \succ a_3 \succ \cdots \succ a_K$.

To evaluate the three bounds, we first identify the stable matching and the regret gaps of the target player p_2.

Lemma 28 (Stable matching and regret gaps)**.** *The market has the unique stable matching $\{(p_1, a_1), (p_2, a_2)\}$. Player p_2 has regret gap $\Delta_{2,2} = 0$ and $\Delta_{2,\ell} = 1/2 - 1/(K\ell) \in [1/4, 1/2)$ for every $\ell \neq 2$.*

Proof. If p_1 does not receive a_1, it prefers a_1 to its assigned arm. Arm a_1 is then either unmatched or held by p_2, whom it ranks below p_1. Thus (p_1, a_1) blocks, so every stable matching must contain this pair. After fixing it, a_2 is unmatched unless it is assigned to p_2. Since p_2 ranks a_2 first, stability also forces (p_2, a_2). The resulting matching gives both players their first choices and is therefore stable. In particular, $\underline{m}(2) = 2$ and $\mu_{2,\underline{m}(2)} = 1/2$. For $\ell \neq 2$, the regret gap and the gap of triplet (p_2, a_2, a_ℓ) satisfy

$$\Delta_{2,\ell} = \mu_{2,\underline{m}(2)} - \mu_{2,\ell} = \mu_{2,2} - \mu_{2,\ell} = \Delta_{2,2,\ell} = \frac{1}{2} - \frac{1}{K\ell}. \tag{17}$$

Since $K \geq 4$ and $\ell \geq 1$, we have $0 < 1/(K\ell) \leq 1/4$, giving the stated interval. Consequently, $0 \leq \Delta_{2,m(2)} \leq 1/2$ for every $m \in \mathcal{M}$. □

G.2 The Bound Λ_2

The arms with positive regret gap are exactly a_ℓ with $\ell \neq 2$, so (1) becomes

$$\Lambda_2(T) = \sum_{\ell \neq 2} \Delta_{2,\ell} \min_{Q \in \mathcal{C}(\mathcal{M}_{2,\ell})} \sum_{(p_j, a_k, a_{k'}) \in Q} \left(5 + \frac{6 \log T}{\Delta_{j,k,k'}^2}\right).$$

For each $\ell \neq 2$, the set $\mathcal{M}_{2,\ell}$ contains every matching that assigns a_ℓ to p_2, with p_1 assigned any other arm. All these matchings are unstable by Lemma 28. The common triplet (p_1, a_1, a_2) has gap $\Delta_{1,1,2} = 1/K$, so its count bound is $5 + 6K^2 \log T = \Theta(K^2 \log T)$. For each $\ell \neq 2$, let $m_\ell = \{(p_1, a_2), (p_2, a_\ell)\}$.

To lower bound the cost of any cover of $\mathcal{M}_{2,\ell}$, we characterize all blocking triplets of m_ℓ and their gaps.

Lemma 29. *For every $\ell \neq 2$, the complete set of blocking triplets of m_ℓ is $\{(p_1, a_1, a_2)\} \cup \{(p_2, a_j, a_\ell) : 1 \leq j < \ell,\ j \neq 2\}$. Every such triplet has gap at most $1/K$, and m_1 is blocked only by (p_1, a_1, a_2).*

Proof. Player p_1 prefers only a_1 to its current arm a_2. If $\ell = 1$, arm a_1 holds p_2 and prefers p_1. If $\ell \geq 3$, arm a_1 is unmatched. Thus (p_1, a_1, a_2) is the only blocking triplet of p_1 in either case.

Player p_2 cannot form a blocking pair with a_2, which prefers its holder p_1. Among the remaining arms, it prefers a_j to a_ℓ exactly when $j < \ell$ and $j \neq 2$. Every such arm is unmatched, giving precisely the triplets (p_2, a_j, a_ℓ) in the statement. Their gaps are

$$\Delta_{2,j,\ell} = \mu_{2,j} - \mu_{2,\ell} = \frac{1}{Kj} - \frac{1}{K\ell} = \frac{1}{K}\left(\frac{1}{j} - \frac{1}{\ell}\right) \in (0, 1/K).$$

There are no such arms when $\ell = 1$, proving the claim about m_1. □

Lower bound. Fix $\ell \neq 2$ and any cover $Q \in \mathcal{C}(\mathcal{M}_{2,\ell})$. Since $m_\ell \in \mathcal{M}_{2,\ell}$, at least one triplet in Q must block m_ℓ. Lemma 29 gives a gap of at most $1/K$ for every such triplet, so its count bound is at least $5 + 6K^2 \log T$. Taking the minimum over covers gives

$$\min_{Q \in \mathcal{C}(\mathcal{M}_{2,\ell})} \sum_{(p_j, a_k, a_{k'}) \in Q} \left(5 + \frac{6 \log T}{\Delta^2_{j,k,k'}}\right) \geq 5 + 6K^2 \log T.$$

Multiplying by the regret gap and summing over the $K - 1$ values of $\ell \neq 2$ gives

$$\begin{aligned} \Lambda_2(T) &\geq \sum_{\ell \neq 2} \Delta_{2,\ell} \left(5 + 6K^2 \log T\right) \\ &\geq \frac{K-1}{4} \left(5 + 6K^2 \log T\right) = \Omega(K^3 \log T), \end{aligned} \tag{18}$$

where the second inequality uses $\Delta_{2,\ell} \geq 1/4$.

Upper bound. For each $\ell \neq 2$, choose $Q_\ell = \{(p_1, a_1, a_2), (p_2, a_2, a_\ell)\}$. To verify this cover, take any $m \in \mathcal{M}_{2,\ell}$, so that $m(2) = \ell$. If $m(1) = 2$, then $m = m_\ell$, which is blocked by (p_1, a_1, a_2) by Lemma 29. If $m(1) \neq 2$, neither player holds a_2, so this arm is unmatched. Player p_2 prefers a_2 to a_ℓ, and hence (p_2, a_2, a_ℓ) blocks m.

Using $\Delta_{2,2,\ell} = \Delta_{2,\ell} \geq 1/4$ from (17), we obtain

$$\begin{aligned} &\min_{Q \in \mathcal{C}(\mathcal{M}_{2,\ell})} \sum_{(p_j, a_k, a_{k'}) \in Q} \left(5 + \frac{6 \log T}{\Delta^2_{j,k,k'}}\right) \\ &\quad \leq \left(5 + 6K^2 \log T\right) + \left(5 + \frac{6 \log T}{\Delta^2_{2,2,\ell}}\right) \\ &\quad = 10 + 6K^2 \log T + \frac{6 \log T}{\Delta^2_{2,\ell}} \leq 10 + 6K^2 \log T + 96 \log T. \end{aligned}$$

Multiplying by $\Delta_{2,\ell} \leq 1/2$ and summing over $\ell \neq 2$ now gives

$$\begin{aligned} \Lambda_2(T) &\leq \sum_{\ell \neq 2} \Delta_{2,\ell} \left(10 + 6K^2 \log T + 96 \log T\right) \\ &\leq \frac{K-1}{2} \left(10 + 6K^2 \log T + 96 \log T\right) = O(K^3 \log T). \end{aligned} \tag{19}$$

Together with (18), this proves the asserted order of $\Lambda_2(T)$.

G.3 The Bounds Γ_2 and Ψ_2

Upper bound. Fix any feasible solution $(x_m)_{m \in \mathcal{M}}$ to (2). Matchings with $m(2) = 2$ contribute zero regret. We separate those with $m(1) = 2$ and group the rest by the arm $m(2) = \ell \neq 2$ assigned to the target. This gives

$$\begin{aligned} \sum_{m \in \mathcal{M}} \Delta_{2,m(2)} x_m &= \sum_{\substack{m \in \mathcal{M} \\ m(1)=2}} \Delta_{2,m(2)} x_m + \sum_{\substack{m \in \mathcal{M} \\ m(1) \neq 2,\ m(2) \neq 2}} \Delta_{2,m(2)} x_m \\ &= \sum_{\substack{m \in \mathcal{M} \\ m(1)=2}} \Delta_{2,m(2)} x_m + \sum_{\ell \neq 2} \Delta_{2,\ell} \sum_{\substack{m \in \mathcal{M} \\ m(1) \neq 2,\ m(2)=\ell}} x_m. \end{aligned} \tag{20}$$

For the first group, every matching with $m(1) = 2$ is some m_ℓ and is blocked by (p_1, a_1, a_2). Conversely, membership in $B_{1,1,2}$ requires $m(1) = 2$ by the definition of a blocking triplet. The corresponding triplet constraint therefore gives

$$\sum_{\substack{m \in \mathcal{M} \\ m(1)=2}} x_m = \sum_{m \in B_{1,1,2}} x_m \leq 5 + \frac{6 \log T}{\Delta^2_{1,1,2}} = 5 + 6K^2 \log T.$$

Since $x_m \geq 0$ and $\Delta_{2,m(2)} \leq 1/2$ for every matching by Lemma 28, the first contribution satisfies

$$\sum_{\substack{m\in\mathcal{M}\\ m(1)=2}} \Delta_{2,m(2)} x_m \leq \frac{1}{2} \sum_{\substack{m\in\mathcal{M}\\ m(1)=2}} x_m \leq \frac{1}{2}\left(5 + 6K^2 \log T\right). \tag{21}$$

For the second group, fix $\ell \neq 2$. Every matching with $m(1) \neq 2$ and $m(2) = \ell$ leaves a_2 unmatched, so it belongs to $B_{2,2,\ell}$. Since $x_m \geq 0$, the triplet constraint and (17) give

$$\begin{aligned}\sum_{\substack{m\in\mathcal{M}\\ m(1)\neq 2,\ m(2)=\ell}} x_m &\leq \sum_{m\in B_{2,2,\ell}} x_m \\ &\leq 5 + \frac{6\log T}{\Delta_{2,2,\ell}^2} = 5 + \frac{6\log T}{\Delta_{2,\ell}^2}.\end{aligned}$$

Multiplying by the positive regret gap of this group yields

$$\begin{aligned}\Delta_{2,\ell} \sum_{\substack{m\in\mathcal{M}\\ m(1)\neq 2,\ m(2)=\ell}} x_m &\leq \Delta_{2,\ell}\left(5 + \frac{6\log T}{\Delta_{2,\ell}^2}\right) \\ &= 5\Delta_{2,\ell} + \frac{6\log T}{\Delta_{2,\ell}} \\ &\leq \frac{5}{2} + 24\log T.\end{aligned}$$

The last inequality uses $\Delta_{2,\ell} \leq 1/2$ in the first term and $\Delta_{2,\ell} \geq 1/4$ in the second. Summing these group bounds over the $K-1$ values of $\ell \neq 2$ gives

$$\sum_{\ell\neq 2} \Delta_{2,\ell} \sum_{\substack{m\in\mathcal{M}\\ m(1)\neq 2,\ m(2)=\ell}} x_m \leq (K-1)\left(\frac{5}{2} + 24\log T\right). \tag{22}$$

Substituting (21) and (22) into (20) bounds the objective of every feasible solution. Taking the maximum therefore gives

$$\Gamma_2(T) \leq \frac{1}{2}\left(5 + 6K^2\log T\right) + (K-1)\left(\frac{5}{2} + 24\log T\right) = O(K^2 \log T). \tag{23}$$

The horizon constraint also gives $\sum_{m\in\mathcal{M}} \Delta_{2,m(2)} x_m \leq \frac{1}{2}\sum_{m\in\mathcal{M}} x_m \leq T/2$ for every feasible solution. Maximizing gives $\Gamma_2(T) \leq T/2$ and hence $\Gamma_2(T) = O(\min\{T, K^2 \log T\})$.

Lower bounds. Since $\Psi_2(T) \leq \Gamma_2(T)$, a feasible solution for $\Psi_2(T)$ gives a lower bound on both programs. Consider the matching $m_1 = \{(p_1, a_2), (p_2, a_1)\}$ and set $x_{m_1} = \min\{T, 6K^2 \log T\}$ and $x_m = 0$ for $m \in \mathcal{M} \setminus \{m_1\}$. All variables are nonnegative, and $\sum_{m\in\mathcal{M}} x_m = x_{m_1} \leq T$ satisfies the horizon constraint. By Lemma 29, m_1 is blocked only by (p_1, a_1, a_2). The constraint for this triplet is satisfied because

$$\sum_{m\in B_{1,1,2}} x_m = x_{m_1} \leq 5 + 6K^2\log T = 5 + \frac{6\log T}{\Delta_{1,1,2}^2}.$$

Every other triplet constraint has left-hand side zero. If $m_1 \in \mathcal{S}_{j,k'}(d)$, then some triplet $(p_j, a_k, a_{k'})$ with $\Delta_{j,k,k'} \geq d$ must block m_1. The only such triplet is (p_1, a_1, a_2), so the assigned pair is (p_1, a_2) and the threshold satisfies $0 < d \leq 1/K$. For every such threshold,

$$\left\lfloor \frac{6\log T}{d^2} \right\rfloor + 2 + 2\sqrt{H_{1,2}(d) + 1} \geq \frac{6\log T}{d^2} \geq 6K^2 \log T.$$

Taking the minimum with T gives $C_{1,2}(d, T) \geq x_{m_1}$. Every joint constraint whose set excludes m_1 has left-hand side zero. Thus the chosen values satisfy every constraint of $\Psi_2(T)$. Only m_1 contributes to its objective, giving

$$\Gamma_2(T) \geq \Psi_2(T) \geq \Delta_{2,1} x_{m_1} = \left(\frac{1}{2} - \frac{1}{K}\right)\min\{T, 6K^2\log T\} \geq \frac{1}{4}\min\{T, K^2 \log T\}. \tag{24}$$

The last inequality uses $1/2 - 1/K \geq 1/4$ and $6K^2 \log T \geq K^2 \log T$. The upper bound on $\Gamma_2(T)$ also bounds $\Psi_2(T)$, proving that both are $\Theta(\min\{T, K^2 \log T\})$ with constants independent of K and T.

Comparison of the bounds. We now verify that the improvement by a factor of order K persists even after capping $\Lambda_2(T)$ by T. If $T \geq K^3 \log T$, then $T \geq K^2 \log T$. The preceding bounds therefore give $\Gamma_2(T) = \Theta(K^2 \log T)$ and $\Lambda_2(T) = \Theta(K^3 \log T)$, so $\Lambda_2(T)/\Gamma_2(T) = \Theta(K)$. In this regime, $\Lambda_2(T) = O(T)$, so the separation by a factor of K persists after capping $\Lambda_2(T)$ by T.

H Sharing Observation Counts

H.1 Market Construction

The market M_K^{B} has $K - 1$ players and K arms, where $K \geq 3$. Let $\zeta = 1/(4\sqrt{K})$ and take independent Bernoulli rewards. Player p_1 has means

$$\mu_{1,K} = \frac{1}{2}, \qquad \mu_{1,k} = \frac{1}{2} + \zeta\left(1 - \frac{k-1}{K^3}\right) \quad (1 \leq k < K). \tag{25}$$

Every other player p_j, with $2 \leq j < K$, has means

$$\mu_{j,j} = \frac{1}{2} + \zeta, \qquad \mu_{j,k} = \frac{1}{2} + \frac{\zeta}{k+1} \quad (k \neq j). \tag{26}$$

All means lie in $(0, 1)$ and are distinct for each player. Player p_1 ranks the arms by increasing index. Every other player ranks its own arm first and the remaining arms by increasing index. Every arm a_k with $k < K$ ranks $p_2 \succ \cdots \succ p_{K-1} \succ p_1$, while a_K ranks $p_1 \succ p_2 \succ \cdots \succ p_{K-1}$. Every matching assigns distinct arms to all $K - 1$ players and leaves exactly one arm unmatched.

We first identify the stable matching of this market.

Lemma 30. *The market has the unique stable matching that assigns a_j to p_j for every $1 \leq j < K$ and leaves a_K unmatched.*

Proof. Let m be any stable matching. The pair (p_2, a_2) consists of mutual first choices, so $m(2) = 2$. For $3 \leq j < K$, suppose that $m(h) = h$ for every $2 \leq h < j$. If $m(j) \neq j$, then a_j is either unmatched or held by a player in $\{p_1, p_{j+1}, \ldots, p_{K-1}\}$, since the preceding players already hold their own arms. Player p_j ranks a_j first, and a_j ranks p_j above every such holder, so (p_j, a_j) would block in either case. Thus $m(j) = j$, and induction fixes all pairs with $2 \leq j < K$. The remaining player is p_1, and the remaining arms are a_1, a_K. If p_1 receives a_K, then a_1 is unmatched and (p_1, a_1) blocks. Hence $m(1) = 1$. This matching gives every player its first choice and is therefore stable, completing the uniqueness argument. □

Since the stable matching assigns a_1 to p_1, its regret gaps are nonnegative and satisfy $\Delta_{1,1} = 0$, $\Delta_{1,K} = \zeta$, and $\Delta_{1,j} = (j-1)\zeta/K^3$ for $2 \leq j < K$. To control the contribution of arms $a_2, \ldots, a_{K-1}$ to the regret bounds, define

$$s_K := \sum_{j=2}^{K-1} \Delta_{1,j} = \frac{(K-2)(K-1)\zeta}{2K^3} \leq \frac{\zeta}{2K}. \tag{27}$$

H.2 Blocking Triplets of the Matchings m_j

The matching m_1 assigns a_K to p_1 and a_j to every p_j with $2 \leq j < K$, leaving a_1 unmatched. For $2 \leq j < K$, matching m_j assigns a_K to p_1 and a_1 to p_j, retaining all other stable pairs and leaving a_j unmatched. Let $\tau_j = (p_1, a_j, a_K)$ for $1 \leq j < K$. The following lemma characterizes the blocking triplets of these matchings and establishes their pairwise disjointness, which we use to lower bound $\Lambda_1(T)$ and $\Gamma_1(T)$.

Lemma 31. *The only blocking triplet of m_1 is τ_1. For $2 \leq j < K$, the only blocking triplets of m_j are τ_j and (p_j, a_j, a_1). These sets are pairwise disjoint, and every triplet in them has gap in $[\zeta/2, \zeta]$.*

Proof. In m_1, every player p_j with $2 \le j < K$ receives its first choice. Player p_1 prefers every other arm to a_K. Arm a_1 is unmatched, giving τ_1. Each remaining arm a_k with $2 \le k < K$ ranks its holder p_k above p_1. Thus τ_1 is the unique blocking triplet.

In m_j with $2 \le j < K$, every player except p_1, p_j receives its first choice. Player p_j receives a_1 and prefers only a_j. Since a_j is unmatched, this gives (p_j, a_j, a_1). Player p_1 also forms a blocking pair with the unmatched arm a_j, giving τ_j. Every other arm a_k with $k < K$ is held by a player ranked above p_1, so no other blocking pair involves p_1.

The gap of τ_j is $\zeta(1 - (j-1)/K^3) \in [\zeta/2, \zeta]$, and the gap of (p_j, a_j, a_1) is exactly $\zeta/2$. All the listed triplets are distinct across $m_1, \dots, m_{K-1}$. □

H.3 Evaluation of the Three Bounds

We evaluate the three bounds using covers of the sets $\mathcal{M}_{1,j}$ with $\Delta_{1,j} > 0$. Recall that $\mathcal{M}_{1,j}$ consists of unstable matchings that assign a_j to p_1. For $2 \le j \le K$, every matching with $m(1) = j$ is unstable, since the unique stable matching assigns a_1 to p_1. Thus, for these indices, $\mathcal{M}_{1,j} = \{m \in \mathcal{M} : m(1) = j\}$.

Take any $m \in \mathcal{M}_{1,K}$. This matching assigns a_K to p_1 and leaves exactly one other arm unmatched, say a_j with $j < K$. Player p_1 prefers a_j to a_K, and the unmatched arm accepts p_1. Hence $\tau_j = (p_1, a_j, a_K)$ blocks m, which means $m \in B_{1,j,K}$. The index j can depend on m, but every $m \in \mathcal{M}_{1,K}$ belongs to at least one of these $K-1$ blocking sets. Therefore $\mathcal{M}_{1,K} \subseteq \bigcup_{j=1}^{K-1} B_{1,j,K}$. Conversely, any matching in $B_{1,j,K}$ assigns a_K to p_1 and is unstable. Every such blocking set is nonempty because it contains m_j. The smallest gap among these triplets is $\Delta_{1,K-1,K} = \zeta(1 - (K-2)/K^3) \ge \zeta/2$. Hence

$$\mathcal{M}_{1,K} = \bigcup_{j=1}^{K-1} B_{1,j,K} = \mathcal{S}_{1,K}(\Delta_{1,K-1,K}), \qquad H_{1,K}(\Delta_{1,K-1,K}) = K - 1. \tag{28}$$

Now fix $2 \le j < K$ and take any $m \in \mathcal{M}_{1,j}$. Since p_1 holds a_j, player p_j must receive some $a_{k'} \neq a_j$. Player p_j ranks its own arm a_j first, and a_j ranks p_j above its holder p_1. Thus $(p_j, a_j, a_{k'})$ blocks m, so $m \in B_{j,j,k'}$. As k' ranges over the $K-1$ arms other than a_j, these sets cover $\mathcal{M}_{1,j}$. The corresponding gaps satisfy

$$\Delta_{j,j,k'} = \left(\frac{1}{2} + \zeta\right) - \left(\frac{1}{2} + \frac{\zeta}{k'+1}\right) = \zeta\left(1 - \frac{1}{k'+1}\right) \ge \frac{\zeta}{2}.$$

These sets together with $\mathcal{M}_{1,K}$ contain all matchings with positive target regret gap. Every remaining matching assigns a_1 to p_1 and has zero target regret gap.

A triplet with gap at most ζ has count bound at least $5 + 6\log T/\zeta^2$, while one with gap at least $\zeta/2$ has count bound at most $5 + 24\log T/\zeta^2$. Since $\zeta^{-2} = 16K$, $K \ge 3$, and $T \ge 2$, $\frac{\log T}{\zeta^2} = 16K \log T \ge 48 \log 2 > 5$. Using $5 \le \log T/\zeta^2$ gives

$$\frac{6\log T}{\zeta^2} \le 5 + \frac{6\log T}{\zeta^2} \le \frac{7\log T}{\zeta^2},$$
$$\frac{24\log T}{\zeta^2} \le 5 + \frac{24\log T}{\zeta^2} \le \frac{25\log T}{\zeta^2}.$$

Thus both expressions are $\Theta(\log T/\zeta^2) = \Theta(K \log T)$, with constants independent of K and T.

The bound $\Lambda_1(T)$. *Upper bound.* The cover $\{\tau_1, \dots, \tau_{K-1}\}$ of $\mathcal{M}_{1,K}$ has cost at most $(K-1)(5 + 24\log T/\zeta^2)$. For each $2 \le j < K$, the cover $\{(p_j, a_j, a_{k'}) : k' \in [K] \setminus \{j\}\}$ of $\mathcal{M}_{1,j}$ contains $K-1$ triplets with gaps at least $\zeta/2$, so it has the same cost bound. Weighting these cover costs by the regret gaps gives

$$\Lambda_1(T) \le (K-1)\left(\zeta + \sum_{j=2}^{K-1} \Delta_{1,j}\right)\left(5 + \frac{24\log T}{\zeta^2}\right) = (K-1)(\zeta + s_K)\left(5 + \frac{24\log T}{\zeta^2}\right).$$

Lower bound. All terms in (1) are nonnegative. Retaining only the contribution of $\mathcal{M}_{1,K}$ and using $\Delta_{1,K} = \zeta$ gives

$$\Lambda_1(T) \geq \zeta \min_{Q \in \mathcal{C}(\mathcal{M}_{1,K})} \sum_{(p_r, a_k, a_{k'}) \in Q} \left(5 + \frac{6 \log T}{\Delta^2_{r,k,k'}}\right).$$

Take any $Q \in \mathcal{C}(\mathcal{M}_{1,K})$. The matchings $m_1, \ldots, m_{K-1}$ all belong to $\mathcal{M}_{1,K}$, so the cover must contain a blocking triplet for each. Lemma 31 shows that these matchings have pairwise disjoint sets of blocking triplets, all with gaps at most ζ. The cover therefore needs at least $K-1$ distinct triplets, each with cost at least $5 + 6 \log T/\zeta^2$, giving

$$\sum_{(p_r, a_k, a_{k'}) \in Q} \left(5 + \frac{6 \log T}{\Delta^2_{r,k,k'}}\right) \geq (K-1) \left(5 + \frac{6 \log T}{\zeta^2}\right).$$

Minimizing over $Q \in \mathcal{C}(\mathcal{M}_{1,K})$ and using $\Delta_{1,K} = \zeta$ gives

$$\Lambda_1(T) \geq \zeta(K-1) \left(5 + \frac{6 \log T}{\zeta^2}\right) \geq \frac{6(K-1) \log T}{\zeta}.$$

Combining the two bounds, we have

$$\begin{aligned} \Lambda_1(T) &\geq \zeta(K-1) \left(5 + \frac{6 \log T}{\zeta^2}\right), \\ \Lambda_1(T) &\leq (K-1)(\zeta + s_K) \left(5 + \frac{24 \log T}{\zeta^2}\right). \end{aligned} \tag{29}$$

Using (27), we obtain $\Lambda_1(T) = \Theta(K \log T/\zeta)$.

The bound $\Gamma_1(T)$. *Upper bound.* Fix any feasible solution $(x_m)_{m \in \mathcal{M}}$ to (2). Grouping the objective by the arm assigned to p_1 gives

$$\begin{aligned} \sum_{m \in \mathcal{M}} \Delta_{1,m(1)} x_m &= \sum_{j=1}^{K} \Delta_{1,j} \sum_{\substack{m \in \mathcal{M} \\ m(1)=j}} x_m \\ &= \zeta \sum_{m \in \mathcal{M}_{1,K}} x_m + \sum_{j=2}^{K-1} \frac{(j-1)\zeta}{K^3} \sum_{m \in \mathcal{M}_{1,j}} x_m. \end{aligned} \tag{30}$$

The second line uses $\Delta_{1,K} = \zeta$ and $\Delta_{1,j} = (j-1)\zeta/K^3$ for $2 \leq j < K$. The group with $m(1) = 1$ contributes zero because $\Delta_{1,1} = 0$. For assignments of a_K, use the cover in (28). Every triplet (p_1, a_j, a_K) with $j < K$ has gap at least $\zeta/2$. Since $x_m \geq 0$, the triplet constraints give

$$\sum_{m \in \mathcal{M}_{1,K}} x_m \leq \sum_{j=1}^{K-1} \sum_{m \in B_{1,j,K}} x_m \leq \sum_{j=1}^{K-1} \left(5 + \frac{6 \log T}{\Delta^2_{1,j,K}}\right) \leq (K-1) \left(5 + \frac{24 \log T}{\zeta^2}\right).$$

For assignments of a_j with $2 \leq j < K$, the cover $\mathcal{M}_{1,j} \subseteq \bigcup_{k' \neq j} B_{j,j,k'}$ contains $K-1$ triplets with gaps at least $\zeta/2$. The same argument gives

$$\begin{aligned} \sum_{m \in \mathcal{M}_{1,j}} x_m &\leq \sum_{k' \neq j} \sum_{m \in B_{j,j,k'}} x_m \\ &\leq \sum_{k' \neq j} \left(5 + \frac{6 \log T}{\Delta^2_{j,j,k'}}\right) \\ &\leq (K-1) \left(5 + \frac{24 \log T}{\zeta^2}\right). \end{aligned} \tag{31}$$

Substituting these count bounds into (30) yields

$$\sum_{m\in\mathcal{M}} \Delta_{1,m(1)}x_m \le (K-1)\left(\zeta + \sum_{j=2}^{K-1}\frac{(j-1)\zeta}{K^3}\right)\left(5+\frac{24\log T}{\zeta^2}\right)$$
$$= (K-1)(\zeta+s_K)\left(5+\frac{24\log T}{\zeta^2}\right).$$

By (27), $s_K \le \zeta/(2K)$, so $(K-1)(\zeta+s_K) \le K\zeta$. Using $5+24\log T/\zeta^2 \le 25\log T/\zeta^2$, the objective is therefore at most $25K\log T/\zeta$.

All target regret gaps lie in $[0,\zeta]$, so the horizon constraint also gives

$$\sum_{m\in\mathcal{M}} \Delta_{1,m(1)}x_m \le \zeta \sum_{m\in\mathcal{M}} x_m \le \zeta T. \tag{32}$$

Maximizing over feasible solutions gives

$$\begin{aligned}\Gamma_1(T) &= \max_{x \text{ feasible}} \sum_{m\in\mathcal{M}} \Delta_{1,m(1)}x_m \\ &\le \min\left\{\zeta T, (K-1)(\zeta+s_K)\left(5+\frac{24\log T}{\zeta^2}\right)\right\} \\ &\le \min\left\{\zeta T, \frac{25K\log T}{\zeta}\right\}.\end{aligned}$$

Lower bound. For each m_j with $1 \le j < K$, set $x_{m_j} = \min\{T/(K-1), 5+6\log T/\zeta^2\}$ and set all other variables to zero. These values are nonnegative and satisfy

$$\sum_{m\in\mathcal{M}} x_m = (K-1)\min\left\{\frac{T}{K-1}, 5+\frac{6\log T}{\zeta^2}\right\} = \min\left\{T, (K-1)\left(5+\frac{6\log T}{\zeta^2}\right)\right\} \le T.$$

By Lemma 31, a triplet can block at most one of $m_1,\dots,m_{K-1}$. If it blocks m_j, its constraint has left-hand side x_{m_j}. Its gap is at most ζ, so the right-hand side is at least $5+6\log T/\zeta^2$, which is no smaller than x_{m_j}. Every other triplet constraint has left-hand side zero. Thus all triplet constraints hold, and the chosen values give a feasible solution. Each m_j assigns a_K to p_1, so its regret gap is ζ and the objective is $\zeta\min\left\{T,(K-1)\left(5+\frac{6\log T}{\zeta^2}\right)\right\}$. Combining the lower and upper bounds yields

$$\begin{aligned}\Gamma_1(T) &\ge \zeta\min\left\{T,(K-1)\left(5+\frac{6\log T}{\zeta^2}\right)\right\}, \\ \Gamma_1(T) &\le \min\left\{\zeta T,(K-1)(\zeta+s_K)\left(5+\frac{24\log T}{\zeta^2}\right)\right\}.\end{aligned} \tag{33}$$

Since $5+6\log T/\zeta^2 = \Theta(\log T/\zeta^2)$, the feasible solution has objective

$$\zeta\min\left\{T,(K-1)\left(5+\frac{6\log T}{\zeta^2}\right)\right\} = \Theta\left(\min\left\{\zeta T, \frac{K\log T}{\zeta}\right\}\right).$$

This matches the upper bound and proves $\Gamma_1(T) = \Theta(\min\{\zeta T, K\log T/\zeta\})$.

The bound $\Psi_1(T)$. The joint constraint replaces the sum of separate bounds on $\mathcal{M}_{1,K}$ by one bound on its total count. The triplet count bounds for the other groups with positive regret gaps remain available. Fix any feasible solution to (10). Since $B_{1,K-1,K}$ is nonempty, the joint constraint at $d=\Delta_{1,K-1,K}$ is included in the program. Using (28) and $\Delta_{1,K-1,K} \ge \zeta/2$ gives

$$\sum_{m\in\mathcal{M}_{1,K}} x_m \le C_{1,K}(\Delta_{1,K-1,K},T) \le \min\left\{T, \left\lfloor\frac{24\log T}{\zeta^2}\right\rfloor + 2 + 2\sqrt{K}\right\} =: c_T. \tag{34}$$

The solution also satisfies all triplet constraints, so (31) still bounds $\sum_{m\in\mathcal{M}_{1,j}} x_m$ for every $2 \le j < K$. Substituting these bounds into (30) gives objective at most $\zeta c_T + (K-1)s_K\left(5+\frac{24\log T}{\zeta^2}\right)$. Equation (32) gives the additional upper bound ζT.

For the lower bound, set $x_{m_1} = \min\{T, 6\log T/\zeta^2\}$ and all other variables to zero. These values are nonnegative and satisfy the horizon constraint. The only blocking triplet of m_1 is τ_1, whose gap is ζ and count bound is $5 + 6\log T/\zeta^2$. Its constraint holds because $x_{m_1} \le 5 + 6\log T/\zeta^2$, while every other triplet constraint has left-hand side zero. If $m_1 \in \mathcal{S}_{j,k'}(d)$, then some triplet $(p_j, a_k, a_{k'})$ with $\Delta_{j,k,k'} \ge d$ must block m_1. That triplet can only be τ_1, so the assigned pair must be (p_1, a_K) and the threshold must satisfy $0 < d \le \zeta$. For every such threshold, $\lfloor u \rfloor \ge u - 1$ gives

$$\left\lfloor \frac{6\log T}{d^2} \right\rfloor + 2 + 2\sqrt{H_{1,K}(d) + 1} \ge \frac{6\log T}{d^2} \ge \frac{6\log T}{\zeta^2}.$$

Taking the minimum with T shows that $C_{1,K}(d,T) \ge x_{m_1}$. Every joint constraint whose set does not contain m_1 has left-hand side zero. Thus the chosen values satisfy all constraints of $\Psi_1(T)$ and have objective $\zeta\min\{T, 6\log T/\zeta^2\}$. Together with the upper bounds, this yields

$$\begin{aligned} \Psi_1(T) &\ge \zeta \min\left\{T, \frac{6\log T}{\zeta^2}\right\}, \\ \Psi_1(T) &\le \min\left\{\zeta T, \zeta c_T + (K-1)s_K\left(5 + \frac{24\log T}{\zeta^2}\right)\right\}. \end{aligned} \tag{35}$$

To bound c_T explicitly, use $K \ge 3$, $T \ge 2$, and $\zeta^{-2} = 16K$ to obtain

$$2 + 2\sqrt{K} \le 4K \le 16K\log 2 \le 16K\log T = \frac{\log T}{\zeta^2}.$$

Since $\lfloor u \rfloor \le u$, the definition of c_T in (34) therefore gives

$$c_T \le \min\left\{T, \frac{24\log T}{\zeta^2} + 2 + 2\sqrt{K}\right\} \le \min\left\{T, \frac{25\log T}{\zeta^2}\right\}.$$

For the contribution of the other assignments, (27) gives $(K-1)s_K \le \zeta/2$. Using the earlier inequality $5 + 24\log T/\zeta^2 \le 25\log T/\zeta^2$, we obtain

$$(K-1)s_K\left(5 + \frac{24\log T}{\zeta^2}\right) \le \frac{\zeta}{2}\left(5 + \frac{24\log T}{\zeta^2}\right) \le \frac{25}{2}\frac{\log T}{\zeta}.$$

Substituting these estimates into (35) and rewriting its lower bound yields

$$\min\left\{\zeta T, \frac{6\log T}{\zeta}\right\} \le \Psi_1(T) \le \min\left\{\zeta T, \frac{75}{2}\frac{\log T}{\zeta}\right\}.$$

Thus $\Psi_1(T) = \Theta(\min\{\zeta T, \log T/\zeta\})$, with constants independent of K and T.

Comparison of the bounds. We now verify that $T \ge K^2\log T$ suffices for observation count sharing to improve the preceding bounds by a factor of order N.

Substituting $\zeta = 1/(4\sqrt{K})$ gives the three orders stated in Theorem 13 for every $K \ge 3$ and $T \ge 2$. If $T \ge K^2\log T$, then

$$\frac{\zeta T}{K\log T/\zeta} = \frac{T}{16K^2\log T} \ge \frac{1}{16},$$
$$\frac{\zeta T}{\log T/\zeta} = \frac{T}{16K\log T} \ge \frac{K}{16} \ge \frac{3}{16}.$$

Each minimum is therefore within constant factors of its logarithmic term, giving $\Gamma_1(T) = \Theta(K\log T/\zeta)$ and $\Psi_1(T) = \Theta(\log T/\zeta)$. Combining these orders with the bound on $\Lambda_1(T)$ gives $\Gamma_1(T)/\Psi_1(T) = \Theta(K) = \Theta(N)$ and $\Lambda_1(T)/\Psi_1(T) = \Theta(K) = \Theta(N)$, since $N = K - 1$.

I Regret Lower Bounds for the Two Market Families

This appendix proves Theorems 12 and 14. Changing the reward distribution of one pair gives a lower bound on how often an algorithm must observe that pair. We derive asymptotic and finite-time versions of this bound, then apply them to both families. In each base market and its alternative, p_1's stable arm is its most preferred arm. For UCB-GS, Theorem 8 therefore gives $0 \le R_1(T) \le O(\log T)$ for fixed K, which implies the regret conditions in Theorems 12 and 14.

I.1 A Common Change-of-Measure Argument

Let $D_{\mathrm{KL}}(P\|Q)$ denote the Kullback–Leibler (KL) divergence between probability distributions P and Q. For Bernoulli distributions with means u and v, write $\mathrm{kl}(u\|v) = u\log(u/v) + (1-u)\log((1-u)/(1-v))$. For a market θ, let $\mu^{\theta}_{j,k}$ denote the mean reward of player-arm pair (p_j, a_k). Let $\mathbb{E}_\theta$ and $\Pr_\theta$ denote expectation and probability in market θ, respectively.

The following lemma provides the observation lower bound used to establish the asymptotic regret lower bounds for both market families.

Lemma 32 (Asymptotic observation lower bound)**.** *Fix two Bernoulli markets θ and λ with the same arm rankings that differ only in the mean of $(p_j, a_{k'})$, with distinct means $\mu^{\theta}_{j,k'}$ and $\mu^{\lambda}_{j,k'}$ in $(0,1)$. Run the same learning algorithm in both markets. Suppose that, for every $\alpha > 0$, $\mathbb{E}_\theta N_{j,k'}(T) = o(T^\alpha)$ and $\mathbb{E}_\lambda[T - N_{j,k'}(T)] = o(T^\alpha)$.*

Then

$$\liminf_{T\to\infty} \frac{\mathbb{E}_\theta N_{j,k'}(T)}{\log T} \geq \frac{1}{\mathrm{kl}(\mu^{\theta}_{j,k'}\|\mu^{\lambda}_{j,k'})}. \tag{36}$$

Proof. These conditions imply that the following event has different probabilities in the two markets. Divergence decomposition and data processing relate this difference to the expected observation count.

Step 1. An event that distinguishes the two markets. Fix any $\alpha \in (0,1)$ and set $E = \{N_{j,k'}(T) < T/2\}$, $p_T = \Pr_\theta(E)$, and $q_T = \Pr_\lambda(E)$.

Under θ, the complement of E requires at least $T/2$ observations of $(p_j, a_{k'})$. Markov's inequality and $\mathbb{E}_\theta N_{j,k'}(T) = o(T^\alpha)$ give

$$1 - p_T = \Pr_\theta\{N_{j,k'}(T) \geq T/2\} \leq \frac{2\mathbb{E}_\theta N_{j,k'}(T)}{T} = o(T^{\alpha-1}).$$

Since $\alpha < 1$, this implies $p_T \to 1$. Under λ, the event E requires more than $T/2$ rounds without an observation of $(p_j, a_{k'})$. Similarly, $\mathbb{E}_\lambda[T - N_{j,k'}(T)] = o(T^\alpha)$ gives

$$q_T = \Pr_\lambda\{T - N_{j,k'}(T) > T/2\} \leq \frac{2\mathbb{E}_\lambda[T - N_{j,k'}(T)]}{T} = o(T^{\alpha-1}).$$

Step 2. Divergence decomposition. Write $\kappa = \mathrm{kl}(\mu^{\theta}_{j,k'}\|\mu^{\lambda}_{j,k'})$. Let $\mathbf{X}_t = (X_1(t), \ldots, X_N(t))$ be the observed reward vector and let $H_t = (m_s, \mathbf{X}_s)_{s=1}^t$ be the complete observed history. Denote the laws of H_T under the two markets by P^T_θ and P^T_λ. Since the policy and arm rankings are the same in both markets, the conditional law of the selected matching given the observed history is also the same. Conditional on a matching m, the reward vector has independent Bernoulli coordinates, and the two markets differ only in the coordinate for $(p_j, a_{k'})$. Its KL divergence is therefore $\kappa\,\mathbf{1}_{\{m(j)=k'\}}$. Applying the divergence decomposition of Lattimore & Szepesvári (2020, Lemma 15.1) to matching actions and their observed reward vectors gives

$$\begin{aligned} D_{\mathrm{KL}}(P^T_\theta\|P^T_\lambda) &= \sum_{t=1}^{T} \mathbb{E}_\theta\left[\sum_{\ell=1}^{N} \mathrm{kl}\Big(\mu^{\theta}_{\ell,m_t(\ell)}\Big\|\mu^{\lambda}_{\ell,m_t(\ell)}\Big)\right] \\ &= \sum_{t=1}^{T} \mathbb{E}_\theta\left[\kappa\,\mathbf{1}_{\{m_t(j)=k'\}}\right] = \kappa\,\mathbb{E}_\theta N_{j,k'}(T). \end{aligned} \tag{37}$$

Step 3. Data processing for the event. The complete history determines $N_{j,k'}(T)$ and hence the indicator $\mathbf{1}_{\{E\}}$. This indicator is Bernoulli with mean p_T under θ and q_T under λ. Applying the data-processing inequality (Lattimore & Szepesvári, 2020, Exercise 14.10) to the map $H_T \mapsto \mathbf{1}_{\{E\}}$ gives

$$\kappa\,\mathbb{E}_\theta N_{j,k'}(T) = D_{\mathrm{KL}}(P^T_\theta\|P^T_\lambda) \geq \mathrm{kl}(p_T\|q_T).$$

To lower bound the binary KL divergence, expand it as

$$\begin{aligned}\mathrm{kl}(p_T\|q_T) &= p_T \log\frac{1}{q_T} + (1-p_T)\log\frac{1}{1-q_T} \quad + p_T\log p_T + (1-p_T)\log(1-p_T)\\ &\geq p_T\log\frac{1}{q_T} - \log 2.\end{aligned}$$

The second term in the expansion is nonnegative, and the last two terms are the negative binary entropy, which is at least $-\log 2$.

Step 4. A logarithmic lower bound on the observation count. From Step 1, $q_T = o(T^{\alpha-1})$, so $q_T \leq T^{\alpha-1}$ for all sufficiently large T. It follows that $\log(1/q_T) \geq (1-\alpha)\log T$ and hence

$$\frac{\kappa\,\mathbb{E}_\theta N_{j,k'}(T)}{\log T} \geq p_T(1-\alpha) - \frac{\log 2}{\log T}.$$

Since $p_T \to 1$, taking the lower limit gives

$$\liminf_{T\to\infty} \frac{\kappa\,\mathbb{E}_\theta N_{j,k'}(T)}{\log T} \geq 1-\alpha.$$

Since this holds for every $\alpha \in (0,1)$, the left-hand side is at least one. Dividing by the fixed positive number κ proves the lemma. □

For the finite-time UCB-GS results, we replace the asymptotic count assumptions by explicit bounds on the same event's probabilities in the two markets.

Lemma 33 (Finite-time observation lower bound)**.** *Consider two Bernoulli markets θ and λ with N players and K arms that differ only in the mean of $(p_j, a_{k'})$, with both values in $(0,1)$. Run UCB-GS with the same arm rankings and the same rule for breaking ties in both markets. Suppose that $a_{k'}$ ranks p_j first and is p_j's first choice in λ. Write $\delta_k = \mu^\lambda_{j,k'} - \mu^\lambda_{j,k} > 0$ for $k \neq k'$ and $\kappa = \mathrm{kl}(\mu^\theta_{j,k'}\|\mu^\lambda_{j,k'})$. If the number of rounds T satisfies*

$$T \geq 64K^2, \qquad \mathbb{E}_\theta N_{j,k'}(T) \leq \frac{T}{8}, \qquad \sum_{k\neq k'} \left\lceil \frac{6\log T}{\delta_k^2} \right\rceil \leq \frac{T}{4}, \tag{38}$$

then

$$\kappa\,\mathbb{E}_\theta N_{j,k'}(T) \geq \frac{1}{4}\log T. \tag{39}$$

Proof. We use the same event $E = \{N_{j,k'}(T) < T/2\}$ as in the asymptotic argument and write $p = \Pr_\theta(E)$ and $q = \Pr_\lambda(E)$.

The count assumption under θ bounds p from below. Under λ, concentration of the empirical means forces at least $T/2$ assignments of $a_{k'}$ to p_j, giving an upper bound on q. We combine these estimates through the same KL argument.

Step 1. The event has large probability in the base market. Markov's inequality and the second condition in (38) give

$$p = 1 - \Pr_\theta\{N_{j,k'}(T) \geq T/2\} \geq 1 - \frac{\mathbb{E}_\theta N_{j,k'}(T)}{T/2} \geq 1 - \frac{T/8}{T/2} = \frac{3}{4}.$$

Step 2. Simultaneous control of the empirical means in the alternative market. For the analysis under λ, generate an independent Bernoulli reward sequence for each pair (p_j, a_k) and reveal its next term whenever that pair is matched. This construction has the same distribution of observations as the online model. Let $\widehat{\mu}_{j,k,n}$ denote the mean of the first n terms of this sequence. For each fixed k and $n \geq 1$, Hoeffding's inequality (Hoeffding, 1963, Theorem 1) gives

$$\Pr_\lambda\left\{|\widehat{\mu}_{j,k,n} - \mu^\lambda_{j,k}| > \sqrt{\frac{\log T}{n}}\right\} \leq 2\exp\left(-2n\frac{\log T}{n}\right) = 2\exp(-2\log T) = \frac{2}{T^2}.$$

Define the event

$$\mathcal{G}_T = \left\{ |\widehat{\mu}_{j,k,n} - \mu^{\lambda}_{j,k}| \le \sqrt{\frac{\log T}{n}} \text{ for all } k \in [K],\ 1 \le n \le T \right\}.$$

There are K choices of arm and T possible positive observation counts. A union bound therefore gives

$$\Pr_{\lambda}(\mathcal{G}_T^c) \le \sum_{k=1}^{K} \sum_{n=1}^{T} \frac{2}{T^2} = \frac{2K}{T}. \tag{40}$$

Thus, on $\mathcal{G}_T$, the same deviation bound holds at the adaptively chosen count $n = N_{j,k}(t-1)$ whenever it is positive.

Step 3. The UCB index of the most preferred arm is at least its mean in the final three quarters of the rounds. Work on $\mathcal{G}_T$ and consider a round $t \in \{\lfloor T/4 \rfloor + 1, \dots, T\}$. Since $T \ge 64K^2 \ge 64$,

$$\frac{3}{2}\log(T/4) - \log T = \frac{1}{2}\log T - \frac{3}{2}\log 4 = \frac{1}{2}\log(T/64) \ge 0.$$

Together with $t > T/4$, this gives $\frac{3}{2}\log t \ge \frac{3}{2}\log(T/4) \ge \log T$. If $n = N_{j,k'}(t-1) > 0$, the UCB index therefore satisfies

$$U_{j,k'}(t) = \widehat{\mu}_{j,k',n} + \sqrt{\frac{3\log t}{2n}} \ge \mu^{\lambda}_{j,k'} - \sqrt{\frac{\log T}{n}} + \sqrt{\frac{3\log t}{2n}} \ge \mu^{\lambda}_{j,k'}.$$

If $N_{j,k'}(t-1) = 0$, the index is $+\infty$, so $U_{j,k'}(t) \ge \mu^{\lambda}_{j,k'}$ still holds.

Step 4. Each other arm can be assigned only a bounded number of times. Continue to work on $\mathcal{G}_T$ in the same interval of rounds. If p_j receives $a_k \ne a_{k'}$, stability under the submitted rankings gives $U_{j,k}(t) \ge U_{j,k'}(t)$.

Indeed, if $U_{j,k}(t) < U_{j,k'}(t)$, then p_j would prefer $a_{k'}$ to a_k under the submitted ranking. Since $a_{k'}$ ranks p_j first, it would accept p_j whether unmatched or assigned to another player. Thus $(p_j, a_{k'})$ would block the selected matching.

For $n = N_{j,k}(t-1) > 0$, the upper deviation bound on $\mathcal{G}_T$ and $t \le T$ give

$$U_{j,k}(t) = \widehat{\mu}_{j,k,n} + \sqrt{\frac{3\log t}{2n}} \le \mu^{\lambda}_{j,k} + \sqrt{\frac{\log T}{n}} + \sqrt{\frac{3\log T}{2n}} = \mu^{\lambda}_{j,k} + \left(1 + \sqrt{\frac{3}{2}}\right)\sqrt{\frac{\log T}{n}}.$$

Suppose that $n \ge 6\log T/\delta_k^2$. Then $\sqrt{\log T/n} \le \delta_k/\sqrt{6}$, and hence

$$U_{j,k}(t) \le \mu^{\lambda}_{j,k} + \frac{1+\sqrt{3/2}}{\sqrt{6}}\,\delta_k < \mu^{\lambda}_{j,k} + \delta_k = \mu^{\lambda}_{j,k'} \le U_{j,k'}(t),$$

where the second inequality uses $(1+\sqrt{3/2})^2 = 5/2 + \sqrt{6} < 6$. This contradicts the index comparison that must hold when p_j is assigned a_k. Thus every such assignment must have prior count $N_{j,k}(t-1) < \frac{6\log T}{\delta_k^2}$. This also holds when the prior count is zero. Since each assignment increases the count, at most $\lceil 6\log T/\delta_k^2 \rceil$ assignments can occur before the threshold, giving

$$\sum_{t=\lfloor T/4 \rfloor + 1}^{T} \mathbf{1}_{\{m_t(j)=k\}} \le \left\lceil \frac{6\log T}{\delta_k^2} \right\rceil.$$

Observations collected before this interval can only reduce the number of additional assignments before the threshold is reached.

Step 5. The event has small probability in the alternative market. There are $T - \lfloor T/4 \rfloor \ge 3T/4$ rounds in this interval. In each round, p_j receives exactly one arm. Summing the bounds from Step 4

over $k \neq k'$ and using the last condition in (38) gives

$$\begin{aligned} N_{j,k'}(T) &\geq \sum_{t=\lfloor T/4 \rfloor + 1}^{T} \mathbf{1}_{\{m_t(j)=k'\}} \\ &= T - \left\lfloor \frac{T}{4} \right\rfloor - \sum_{k \neq k'} \sum_{t=\lfloor T/4 \rfloor + 1}^{T} \mathbf{1}_{\{m_t(j)=k\}} \\ &\geq T - \left\lfloor \frac{T}{4} \right\rfloor - \sum_{k \neq k'} \left\lceil \frac{6 \log T}{\delta_k^2} \right\rceil \\ &\geq \frac{3T}{4} - \frac{T}{4} = \frac{T}{2}. \end{aligned}$$

Thus $E \subseteq \mathcal{G}_T^c$ under λ, and (40) yields $q = \Pr_\lambda(E) \leq \Pr_\lambda(\mathcal{G}_T^c) \leq \frac{2K}{T}$.

Step 6. The probability difference requires logarithmically many observations. The divergence decomposition (37), data processing for E, and the binary entropy bound used above give

$$\begin{aligned} \kappa \, \mathbb{E}_\theta N_{j,k'}(T) &\geq \mathrm{kl}(p \| q) \\ &\geq p \log(1/q) - \log 2 \\ &\geq \frac{1}{2} \log \frac{T}{2K} - \log 2 \\ &= \frac{1}{2} \log \frac{T}{8K} \\ &\geq \frac{1}{4} \log T. \end{aligned}$$

The third inequality uses $p \geq 3/4 \geq 1/2$ and $q \leq 2K/T < 1$. The last inequality follows from $T \geq 64K^2$, which implies $T/(8K) \geq \sqrt{T}$. This proves (39). □

I.2 Regret on the First Market Family

We prove Theorem 12, starting with its asymptotic statement. Fix $K \geq 4$, write $\theta = M_K^{\mathrm{A}}$ and $\lambda = M_K^{\mathrm{A}\prime}$, and consider any learning algorithm π satisfying the theorem's regret conditions. We establish the sharper bound

$$\liminf_{T \to \infty} \frac{R_2^\pi(T; M_K^{\mathrm{A}})}{\log T} \geq \frac{\frac{1}{2} - \frac{1}{K}}{\mathrm{kl}\left(\frac{1}{2} - \frac{1}{K} \middle\| \frac{1}{2} + \frac{1}{K}\right)} \geq \frac{K^2}{48}. \tag{41}$$

The argument uses these conditions to lower bound the number of observations of (p_1, a_2), then bounds the regret they impose on p_2.

Step 1. The stable matching in each market. The alternative λ changes only $\mu_{1,2}$ from $1/2 - 1/K$ to $1/2 + 1/K$. The changed mean lies in $(0, 1)$, and all player rankings remain strict. In θ, the unique stable matching is $\{(p_1, a_1), (p_2, a_2)\}$. In λ, player p_1 ranks a_2 first, and every arm still ranks p_1 first. The pair (p_1, a_2) is therefore present in every stable matching. Among the remaining arms, p_2 prefers a_1, so the unique stable matching is $\{(p_1, a_2), (p_2, a_1)\}$. In particular, p_1's stable arm is its most preferred arm in each market. All of its regret gaps are therefore nonnegative in both markets.

Step 2. From regret to observation counts. Recall that $N_{1,2}(T) = \sum_{t=1}^{T} \mathbf{1}_{\{m_t(1)=2\}}$ and $T - N_{1,2}(T) = \sum_{t=1}^{T} \mathbf{1}_{\{m_t(1)\neq 2\}}$. In θ, assigning a_2 to p_1 incurs regret $1/K$. Discarding the nonnegative contributions from its other assignments gives

$$R_1^\pi(T; \theta) = \mathbb{E}_\theta \sum_{t=1}^{T} \left(\mu_{1,1}^\theta - \mu_{1,m_t(1)}^\theta\right) \geq \frac{1}{K} \mathbb{E}_\theta \sum_{t=1}^{T} \mathbf{1}_{\{m_t(1)=2\}} = \frac{1}{K} \mathbb{E}_\theta N_{1,2}(T).$$

In λ, the mean of a_2 for p_1 is $1/2 + 1/K$, whereas every other arm has mean at most $1/2$. Every assignment other than a_2 thus incurs regret at least $1/K$, and

$$R_1^\pi(T; \lambda) = \mathbb{E}_\lambda \sum_{t=1}^{T} \left(\mu_{1,2}^\lambda - \mu_{1,m_t(1)}^\lambda\right) \geq \frac{1}{K} \mathbb{E}_\lambda \sum_{t=1}^{T} \mathbf{1}_{\{m_t(1)\neq 2\}} = \frac{1}{K} \mathbb{E}_\lambda [T - N_{1,2}(T)].$$

Since K is fixed, the regret conditions imply, for every $\alpha > 0$,

$$0 \le \frac{\mathbb{E}_\theta N_{1,2}(T)}{T^\alpha} \le K \frac{R_1^\pi(T;\theta)}{T^\alpha} \longrightarrow 0,$$
$$0 \le \frac{\mathbb{E}_\lambda[T - N_{1,2}(T)]}{T^\alpha} \le K \frac{R_1^\pi(T;\lambda)}{T^\alpha} \longrightarrow 0.$$

Step 3. The required observations of (p_1, a_2). The two markets differ only in this pair, so the conditions of Lemma 32 now hold with $(j, k') = (1, 2)$. Write the KL divergence of its two reward distributions as $\kappa_K = \mathrm{kl}\big(\frac{1}{2} - \frac{1}{K} \big\| \frac{1}{2} + \frac{1}{K}\big) = \frac{2}{K} \log \frac{K+2}{K-2}$. The lemma gives $\liminf_{T\to\infty} \frac{\mathbb{E}_\theta N_{1,2}(T)}{\log T} \ge \frac{1}{\kappa_K}$. For the explicit constant, use

$$\log \frac{1+x}{1-x} = 2 \int_0^x \frac{1}{1-u^2}\, du \le \frac{2x}{1-x^2}.$$

Taking $x = 2/K \le 1/2$ yields

$$\kappa_K \le \frac{2}{K} \cdot \frac{4/K}{1 - 4/K^2} = \frac{8}{K^2(1 - 4/K^2)} \le \frac{32}{3K^2} \le \frac{12}{K^2},$$

where $1 - 4/K^2 \ge 3/4$ because $K \ge 4$.

Step 4. The regret incurred by p_2. In θ, player p_2's stable arm a_2 has mean $1/2$ and is its most preferred arm. Whenever p_1 receives a_2, the matching must assign a different arm to p_2. Every such arm has mean at most $1/K$, so the regret gap for p_2 in that round is at least $1/2 - 1/K$. Its regret gaps in all other rounds are nonnegative. Consequently,

$$\begin{aligned} R_2^\pi(T;\theta) &= \mathbb{E}_\theta \sum_{t=1}^{T} \big(\mu^\theta_{2,2} - \mu^\theta_{2,m_t(2)}\big) \\ &\ge \left(\frac{1}{2} - \frac{1}{K}\right) \mathbb{E}_\theta \sum_{t=1}^{T} \mathbf{1}_{\{m_t(1)=2\}} \\ &= \left(\frac{1}{2} - \frac{1}{K}\right) \mathbb{E}_\theta N_{1,2}(T). \end{aligned}$$

Combining this inequality with Step 3, and using $1/2 - 1/K \ge 1/4$, gives

$$\liminf_{T\to\infty} \frac{R_2^\pi(T;\theta)}{\log T} \ge \frac{1/2 - 1/K}{\kappa_K} \ge \frac{K^2}{48}.$$

This proves (41).

Now we prove the finite-time bound for UCB-GS.

We apply Lemma 33 to the same two markets, with $(j, k') = (1, 2)$. Assume

$$T \ge 64K^2 \log T. \tag{42}$$

This condition implies $\log T > 1$ and $T \ge 64K^2$.

In the base market, a round assigns a_2 to p_1 exactly when (p_1, a_1, a_2) blocks the selected matching. Its gap is $1/K$, so Lemma 2 gives

$$\mathbb{E}_\theta N_{1,2}(T) \le 5 + \frac{6 \log T}{(1/K)^2} = 5 + 6K^2 \log T \le 8K^2 \log T \le \frac{T}{8}. \tag{43}$$

Here $5 \le 2K^2 \log T$, and the final inequality uses (42). This verifies the count condition in the base market.

In the alternative market, a_2 and p_1 are mutual first choices. The reward gap over a_1 is $1/K$, and the gap over every a_k with $k \ge 3$ is at least $1/2$. The sum of the observation thresholds in (38) is therefore at most

$$\begin{aligned} \big\lceil 6K^2 \log T \big\rceil + (K-2)\,\lceil 24 \log T \rceil &\le \big(6K^2 + 24(K-2)\big) \log T + K - 1 \\ &\le 9K^2 \log T + K - 1 \\ &\le 10K^2 \log T \\ &\le \frac{10}{64} T \le \frac{T}{4}. \end{aligned} \tag{44}$$

The second inequality uses $3K^2-24(K-2) = 3(K-4)^2 \geq 0$, and the third uses $K-1 \leq K^2 \log T$. Thus all conditions of Lemma 33 hold. Combining its conclusion with $\kappa_K \leq 12/K^2$ gives

$$\mathbb{E}_\theta N_{1,2}(T) \geq \frac{\log T}{4\kappa_K} \geq \frac{\log T}{4(12/K^2)} = \frac{K^2 \log T}{48}. \tag{45}$$

Applying the regret inequality from Step 4 to UCB-GS and using $1/2 - 1/K \geq 1/4$ gives

$$R_2(T) \geq \left(\frac{1}{2} - \frac{1}{K}\right) \mathbb{E}_\theta N_{1,2}(T) \geq \frac{1}{4} \cdot \frac{K^2 \log T}{48} = \frac{K^2 \log T}{192}. \tag{46}$$

The corresponding upper bound follows from Theorems 6 and 11, which give $R_2(T) \leq \Psi_2(T) \leq \Gamma_2(T) = O(K^2 \log T)$. This proves the finite-time statement of Theorem 12.

I.3 Regret on the Second Market Family

We prove Theorem 14 by the same change-of-measure argument, with observations of (p_1, a_K) now contributing directly to p_1's regret. For fixed $K \geq 3$ and any learning algorithm π satisfying the theorem's regret conditions, we establish the sharper asymptotic bound

$$\liminf_{T\to\infty} \frac{R_1^\pi(T; M_K^{\mathrm{B}})}{\log T} \geq \frac{\zeta}{\mathrm{kl}(1/2 \| 1/2 + 2\zeta)} \geq \frac{\sqrt{K}}{3}. \tag{47}$$

Let $\theta = M_K^{\mathrm{B}}$ and let $\lambda = M_K^{\mathrm{B}\prime}$ be the alternative obtained by changing only $\mu_{1,K}$ from $1/2$ to $1/2 + 2\zeta$. Since $\zeta = 1/(4\sqrt{K})$, the changed mean lies in $(0, 1)$ and exceeds all other means of p_1. Both markets satisfy the model assumptions, and a_K still ranks p_1 first. Thus p_1's stable arm is a_1 in θ and a_K in λ. In each market, this is p_1's most preferred arm. Write $\kappa_K = \mathrm{kl}(1/2 \| 1/2 + 2\zeta)$ for the KL divergence per observation of (p_1, a_K). Since $16\zeta^2 = 1/K$, direct evaluation and the inequality $-\log(1-x) \leq x/(1-x)$ for $0 < x < 1$ give

$$\kappa_K = -\frac{1}{2}\log(1 - 16\zeta^2) = -\frac{1}{2}\log\left(1 - \frac{1}{K}\right) \leq \frac{1}{2(K-1)} \leq \frac{3}{4K} = 12\zeta^2,$$

where the last inequality uses $K \geq 3$.

In θ, assigning a_K to p_1 incurs regret ζ, and its other regret gaps are nonnegative. In λ, every assignment to another arm incurs regret at least ζ. The regret conditions therefore give, for every $\alpha > 0$,

$$\mathbb{E}_\theta N_{1,K}(T) \leq \frac{R_1^\pi(T; \theta)}{\zeta} = o(T^\alpha),$$
$$\mathbb{E}_\lambda[T - N_{1,K}(T)] \leq \frac{R_1^\pi(T; \lambda)}{\zeta} = o(T^\alpha).$$

Lemma 32 implies $\liminf_{T\to\infty} \mathbb{E}_\theta N_{1,K}(T)/\log T \geq 1/\kappa_K$. Since $R_1^\pi(T; \theta) \geq \zeta \mathbb{E}_\theta N_{1,K}(T)$ and ζ is a positive constant for fixed K, we obtain

$$\liminf_{T\to\infty} \frac{R_1^\pi(T; \theta)}{\log T} \geq \zeta \liminf_{T\to\infty} \frac{\mathbb{E}_\theta N_{1,K}(T)}{\log T} \geq \frac{\zeta}{\kappa_K}.$$

Using $\kappa_K \leq 12\zeta^2$,

$$\frac{\zeta}{\kappa_K} \geq \frac{\zeta}{12\zeta^2} = \frac{1}{12\zeta} = \frac{\sqrt{K}}{3}.$$

This proves (47).

We now prove the finite-time regret bound for UCB-GS. We first establish a lower bound on the expected number of observations of (p_1, a_K).

Lemma 34. *Under UCB-GS on $\theta = M_K^{\mathrm{B}}$, for $K \geq 3$ and $T \geq 384K^2 \log T$,*

$$\mathbb{E}_\theta N_{1,K}(T) \geq \frac{\log T}{48\zeta^2}. \tag{48}$$

Proof. In θ, every matching that assigns a_K to p_1 belongs to $\mathcal{S}_{1,K}(\Delta_{1,K-1,K})$ by (28). Conversely, every round with a matching in this set observes (p_1, a_K). The smallest relevant gap satisfies

$$\Delta_{1,K-1,K} = \zeta\left(1 - \frac{K-2}{K^3}\right) \geq \zeta\left(1 - \frac{1}{K^2}\right) \geq \frac{8}{9}\zeta. \tag{49}$$

By (49), we have

$$\begin{aligned} \mathbb{E}_\theta N_{1,K}(T) &\leq C_{1,K}(\Delta_{1,K-1,K}, T) \\ &\leq \frac{6\log T}{(8\zeta/9)^2} + 2 + 2\sqrt{K} \\ &= \frac{243}{2} K \log T + 2 + 2\sqrt{K} \\ &\leq \frac{275}{2} K \log T \leq \frac{T}{8}. \end{aligned} \tag{50}$$

Here $\zeta^{-2} = 16K$ and $2 + 2\sqrt{K} \leq 4K \leq 16K \log T$. The last inequality follows from $T/8 \geq 48K^2 \log T \geq (275/2)K \log T$.

In λ, the gap satisfies $\mu^\lambda_{1,K} - \mu^\lambda_{1,k} = \zeta(1 + (k-1)/K^3) \geq \zeta$ for every $k < K$. Consequently,

$$\begin{aligned} \sum_{k=1}^{K-1} \left\lceil \frac{6\log T}{(\mu^\lambda_{1,K} - \mu^\lambda_{1,k})^2} \right\rceil &\leq (K-1)\left\lceil \frac{6\log T}{\zeta^2} \right\rceil \\ &\leq (K-1)(96K\log T + 1) \\ &\leq 96K^2 \log T \leq \frac{T}{4}. \end{aligned} \tag{51}$$

The third inequality uses $K - 1 \leq 96K \log T$, and the last uses $T \geq 384K^2 \log T$. The conditions also imply $T \geq 384K^2 \log 2 \geq 64K^2$, so Lemma 33 applies to (p_1, a_K). Its conclusion and $\kappa_K \leq 12\zeta^2$ give

$$\mathbb{E}_\theta N_{1,K}(T) \geq \frac{\log T}{4\kappa_K} \geq \frac{\log T}{48\zeta^2},$$

which proves (48). □

Applying the earlier inequality $R^\pi_1(T;\theta) \geq \zeta \mathbb{E}_\theta N_{1,K}(T)$ to UCB-GS and using Lemma 34 gives

$$R_1(T) \geq \zeta \mathbb{E}_\theta N_{1,K}(T) \geq \frac{\log T}{48\zeta}. \tag{52}$$

Theorem 6 and (35) give the corresponding upper bound. Using $\zeta^{-1} = 4\sqrt{K}$, we obtain $R_1(T) \leq \Psi_1(T) = O\left(\frac{\log T}{\zeta}\right) = O(\sqrt{K}\log T)$. Together with (52), this proves the finite-time statement of Theorem 14 with constants independent of K and T.

J Proof of the Many-to-One Market Regret Bound

We first show that the triplet count and joint count arguments continue to apply in many-to-one markets. We then pass to the corresponding seat market, apply the structural cover argument there, and project the resulting blocking pairs back to the original arms. Finally, we use the α-condition to obtain the sharper regret bound.

J.1 Triplet Counts and Observation Count Sharing

Let $\mathcal{M}_c$ be the set of matchings $m : [N] \to [K]$ satisfying $|\{j \mid m(j) = k\}| \leq c_k$ for each k. All pairs are acceptable and total capacity $C = \sum_k c_k$ is at least N, so player-proposing GS assigns every player. Each player-arm pair (p_j, a_k) has an independent reward stream supported on $[0,1]$ with mean $\mu_{j,k}$. The mean depends only on the pair, and each player's means are distinct across arms. A player and an arm form a blocking pair if the player prefers that arm to its assignment and the arm either has a vacancy or prefers the player to a current holder. Let $\underline{m}$ be the player-pessimal

stable matching and retain the notation $\Delta_{i,k'} = \mu_{i,\underline{m}(i)} - \mu_{i,k'}$ for regret gaps. Put $L = \min\{N, K\}$, $D = \min\{N, K-1\}$, and $\Delta = \min_{j\in[N]} \min_{1\le r\le D}(\mu_{j,\sigma_{j,r}} - \mu_{j,\sigma_{j,r+1}})$.

For distinct arms $a_k, a_{k'}$, define $B_{j,k,k'} = \{m \in \mathcal{M}_c \mid m(j) = k',\ (p_j, a_k) \text{ blocks } m\}$. Whenever $B_{j,k,k'}$ is nonempty, the associated triplet gap is $\Delta_{j,k,k'} = \mu_{j,k} - \mu_{j,k'} > 0$. The sets $\mathcal{S}_{j,k'}(d)$ and the counts $H_{j,k'}(d)$ are defined as in Section 4.2 using these triplets. Let $T_m(T)$ count the rounds selecting $m \in \mathcal{M}_c$, and let $\mathcal{M}_{i,\ell}$ contain the unstable matchings in $\mathcal{M}_c$ that assign a_ℓ to p_i. Covers of this set and the collection $\mathcal{C}(\mathcal{M}_{i,\ell})$ are defined as in Section 3. Use (1), (2), and (10) over $\mathcal{M}_c$ to define Λ_i, Γ_i, and Ψ_i.

If (p_j, a_k) blocks the matching in a round in which p_j receives $a_{k'}$, stability under the submitted rankings gives $U_{j,k'}(t) \ge U_{j,k}(t)$. Otherwise the same pair would block under those rankings, since vacancies and arm rankings are unchanged. Every such round increments the observation count $N_{j,k'}$ exactly once. The triplet count argument in Appendix A and the joint count argument in Appendix B therefore apply without change and give

$$\sum_{m\in B_{j,k,k'}} \mathbb{E}T_m(T) \le 5 + \frac{6\log T}{\Delta^2_{j,k,k'}}, \qquad \sum_{m\in \mathcal{S}_{j,k'}(d)} \mathbb{E}T_m(T) \le C_{j,k'}(d,T).$$

Taking $x_m = \mathbb{E}T_m(T)$ gives a feasible solution to both programs. The arguments of Theorems 3 and 6 then give

$$R_i(T) \le \Psi_i(T) \le \Gamma_i(T) \le \Lambda_i(T). \tag{53}$$

J.2 A Corresponding One-to-One Market

Following the construction in Roth & Sotomayor (1989, Section 4.1), replace arm a_k by c_k seats $(k,1), \dots, (k,c_k)$, each with the same ranking of players as a_k. Each player orders seats of different arms according to its original preferences and prefers smaller indices among seats of the same arm. Write $\rho(k,h) = k$ for the index of the original arm. UCB indices, observation counts, and reward gaps remain defined for player-arm pairs in the original many-to-one market.

For any matching $m \in \mathcal{M}_c$, construct a one-to-one matching u by assigning the holders of each arm to its first seats in the order of that arm's ranking, with $u(j)$ denoting the seat assigned to p_j. No player forms a blocking pair with another seat of its assigned arm. For any player p_j, every earlier seat of its assigned arm is held by a player whom the arm ranks above p_j, while p_j prefers its assigned seat to every later seat. A blocking pair (p_j, e) for u therefore uses a seat of a different arm, and $(p_j, a_{\rho(e)})$ blocks m. Conversely, if (p_j, a_k) blocks m, then p_j forms a blocking pair with an unassigned seat of a_k or a seat held by a player whom a_k ranks below p_j. Thus m is stable if and only if u is stable (Roth & Sotomayor, 1989, Lemma 1).

Every stable matching in the one-to-one market assigns the holders of each original arm to seats in this same order. If a seat were unassigned while a later seat was assigned, the earlier seat would block with the later seat's holder. If an arm ranked the holder of a later seat above the holder of an earlier seat, the holder of the later seat would form a blocking pair with the earlier seat. The projection of any stable matching on the seats is therefore stable in the original market by the equivalence above. Thus each stable matching in the original market corresponds to exactly one stable matching on the seats. Let r be the player-pessimal stable matching in the one-to-one market. Replacing each seat assigned by r with its original arm gives $\underline{m}$. Indeed, the projection of r is stable by the correspondence just proved. If some player received a strictly worse original arm in another stable matching, the corresponding matching on seats would give that player a seat strictly worse than $r(j)$, contradicting the definition of r. Thus the projection gives each player its least preferred stable arm and equals $\underline{m}$. Let $S' = \{r(j) \mid j \in [N]\}$, $S = \{a_{\underline{m}(j)} \mid j \in [N]\}$, and $s = |S|$. We call $r(j)$ player p_j's reference seat and $a_{\underline{m}(j)}$ its reference arm. The set S' contains N reference seats, while the reference arms in S satisfy $s \le \min\{N, K\}$.

J.3 Cover of Matchings with Positive Regret Gaps

Fix a target player p_i. For each matching m with $\Delta_{i,m(i)} > 0$, let u be its corresponding one-to-one matching. Place m in the first class if $(p_i, r(i))$ blocks u. Among the remaining matchings, place m in the second class if u has a blocking pair whose player is assigned to a seat in S'. The remaining

matchings form the third class. Denote the three classes by $\mathcal{M}_i^{(1)}$, $\mathcal{M}_i^{(2)}$, and $\mathcal{M}_i^{(3)}$, respectively. We use these classes to construct a cover by blocking triplets in the original market.

The arguments of Lemmas 23, 24, and 25 use only strict preferences, so they apply to the seat market. For a matching m in the third class with seat assignment u, let each seat in S' choose its preferred holder from r and u, using its holder under r if unassigned in u. Lemma 23 gives a stable matching ν_u on S' with $\nu_u(j) \in \{r(j), u(j)\}$ for every player and

$$\nu_u(i) = u(i) \in S'. \tag{54}$$

By Lemma 24, every player p_q assigned outside S' in u forms a blocking pair $(p_q, r(q))$.

If the third class is nonempty, Lemma 25 gives one seat $e \notin S'$ that is assigned in every corresponding matching u. Its proof takes the player join of the restricted stable matchings ν_u and uses a blocking pair of its extension to the full seat market. For each such u, let p_q be the holder of e. Since $u(i) \in S'$ by (54) while $e \notin S'$, we have $q \neq i$. By Lemma 24, $(p_q, r(q))$ blocks u. The ordered assignment of holders to seats rules out blocking pairs within a player's assigned arm, so $\rho(e) \neq \underline{m}(q)$ and $\mu_{q,\underline{m}(q)} > \mu_{q,\rho(e)}$. Projecting these blocking pairs to the original arms therefore gives

$$\mathcal{M}_i^{(3)} \subseteq \bigcup_{\substack{q \in [N]\setminus\{i\} \\ \mu_{q,\underline{m}(q)} > \mu_{q,\rho(e)}}} B_{q,\underline{m}(q),\rho(e)}. \tag{55}$$

This cover uses at most $N-1$ triplets, each involving two distinct original arms and having a positive gap. The original arm $a_{\rho(e)}$ may belong to S even though its seat e lies outside S'. If $C = N$, every seat belongs to S', so the third class is empty.

J.4 A Finite-Time Regret Bound

We now apply the triplet constraints to this cover, with observations and reward gaps defined for the original player-arm pairs. We first note two rank properties. Every stable arm is among its player's first L choices. Moreover, if a player has a blocking pair, then it has one with an arm among its first L choices. When $L = K$, both claims are immediate. It remains to consider $L = N < K$. If a player receives an arm outside its first N choices, at most $N-1$ of those choices have a holder among the other players. An arm among these choices therefore has no holder and, since its capacity is positive, forms a blocking pair with the player. Such an assignment cannot occur in a stable matching, proving the first statement. This also proves the second statement when the player's assignment is outside its first N choices. If instead a player with a blocking pair holds an arm among its first N choices, every arm forming a blocking pair with that player is preferred to its assignment and is therefore also among these choices.

For an assigned pair $(p_j, a_{k'})$, let $\mathcal{A}_{j,k'}^L$ be the indices $k \neq k'$ of arms among the player's first L choices for which $B_{j,k,k'}$ is nonempty. If $a_{k'}$ has rank v, put $q = \min\{v-1, L\} \leq D$. For a preferred arm of rank $r \leq q$, summing the adjacent gaps from rank r through rank q gives a gap of at least $(q - r + 1)\Delta$, since $q + 1 \leq v$ and $q \leq D$. The integers $q - r + 1$ are distinct as the blocking arm varies, so the same argument as in Lemma 22 gives

$$\sum_{k \in \mathcal{A}_{j,k'}^L} \left(5 + \frac{6 \log T}{\Delta_{j,k,k'}^2}\right) \leq 5D + \frac{\pi^2 \log T}{\Delta^2}. \tag{56}$$

Fix any feasible solution $(x_m)_{m \in \mathcal{M}_c}$ for $\Gamma_i(T)$ and discard its nonpositive objective terms. In the first class, fix an arm $a_{k'}$ assigned to the target in this class. The seat pair $(p_i, r(i))$ projects to the blocking triplet $(p_i, a_{\underline{m}(i)}, a_{k'})$, whose gap equals the regret gap $\Delta_{i,k'}$. Since the reference arm is among the target's first L choices, the regret gap satisfies $\Delta_{i,k'} \in [\Delta, 1]$. The triplet constraint yields

$$\sum_{\substack{m \in \mathcal{M}_i^{(1)} \\ m(i) = k'}} \Delta_{i,m(i)} x_m \leq \Delta_{i,k'} \sum_{m \in B_{i,\underline{m}(i),k'}} x_m \leq 5\Delta_{i,k'} + \frac{6 \log T}{\Delta_{i,k'}} \leq 5 + \frac{6 \log T}{\Delta}.$$

Summing over at most $K-1$ assigned arms bounds the first class's contribution by $(K-1)(5 + 6 \log T/\Delta)$.

For the second class, the corresponding blocking pair in the original market has its player assigned to an arm in S. The rank argument above gives a blocking pair involving one of the same player's first L choices. Thus every matching in this class belongs to some $B_{j,k,k'}$ with $a_{k'} \in S$ and $k \in \mathcal{A}^L_{j,k'}$. There are at most Ns such assigned pairs $(p_j, a_{k'})$. Summing the triplet constraints and applying (56) gives

$$\sum_{m \in \mathcal{M}_i^{(2)}} \Delta_{i,m(i)} x_m \le \sum_{j \in [N]} \sum_{k' : a_{k'} \in S} \sum_{k \in \mathcal{A}^L_{j,k'}} \sum_{m \in B_{j,k,k'}} x_m \le Ns \left(5D + \frac{\pi^2 \log T}{\Delta^2} \right).$$

The third class contributes zero if it is empty. Otherwise, use the common seat e and the cover (55). Every triplet in this cover has a positive gap and uses a reference arm as its blocking arm. This arm is among the player's first L choices, so the gap is at least Δ. Since each regret gap is at most one, summing the nonempty triplet constraints gives

$$\sum_{m \in \mathcal{M}_i^{(3)}} \Delta_{i,m(i)} x_m \le \sum_{\substack{q \in [N] \setminus \{i\} \\ \mu_{q,\underline{m}(q)} > \mu_{q,\rho(e)}}} \sum_{m \in B_{q,\underline{m}(q),\rho(e)}} x_m \le (N-1) \left(5 + \frac{6 \log T}{\Delta^2} \right).$$

Adding these contributions bounds the objective of the feasible solution. The positive objective terms also satisfy $\sum_{m : \Delta_{i,m(i)} > 0} \Delta_{i,m(i)} x_m \le \sum_{m \in \mathcal{M}_c} x_m \le T$. Maximizing over feasible solutions therefore gives

$$\begin{aligned} \Gamma_i(T) \le \min \Bigg\{ T, (K-1) \left(5 + \frac{6 \log T}{\Delta} \right) + Ns \left(5D + \frac{\pi^2 \log T}{\Delta^2} \right) \\ + (N-1) \left(5 + \frac{6 \log T}{\Delta^2} \right) \Bigg\}. \end{aligned} \tag{57}$$

For any player, summing the first D adjacent gaps gives $D\Delta \le 1$. Thus all constant terms are absorbed, giving $\Gamma_i(T) = O(K \log T/\Delta + Ns \log T/\Delta^2)$. Since $s \le \min\{N, K\}$, this bound and (53) imply the general regret guarantee in Theorem 10.

J.5 The α-Condition

We use the formulation of the α-condition in Kong et al. (2024, Definition 6.1) on the strict seat preferences defined above. It requires a stable matching with two possibly different orders of its matched pairs. Denote this stable matching by r. Let $(P_1, \dots, P_N)$ and $(Q_1, \dots, Q_N)$ be the permutations of player indices in the two orders.

1. In the left order $(p_{P_1}, f_1), \dots, (p_{P_N}, f_N)$, where $f_h = r(P_h)$, player p_{P_h} prefers f_h to every f_v with $v > h$ and to every seat unassigned under r.
2. In the right order $(p_{Q_1}, e_1), \dots, (p_{Q_N}, e_N)$, where $e_h = r(Q_h)$, seat e_h prefers p_{Q_h} to every p_{Q_v} with $v > h$.

The α-condition implies that r is the unique stable matching in the seat market (Basu et al., 2021, Theorem 2). Thus r is the player-pessimal stable matching, and the correspondence between the two markets shows that its projection is the unique stable matching $\underline{m}$ of the original market.

The right order allows us to trace a player who receives a seat worse than its seat under r to a blocking pair. Consider any one-to-one matching u assigning every player, and start from a player p_{Q_v} who prefers e_v to $u(Q_v)$. Stop if (p_{Q_v}, e_v) blocks u. Otherwise e_v has a holder p_{Q_w} whom it prefers to p_{Q_v}. The right order gives $w < v$. Since r is stable, p_{Q_w} must prefer e_w to $e_v = u(Q_w)$: otherwise (p_{Q_w}, e_v) would block r, because e_v prefers p_{Q_w} to its holder p_{Q_v} under r. We therefore continue from p_{Q_w}. The index in the right order strictly decreases at each step, so this process ends at a blocking pair involving a player and its seat under r.

Now fix $i = Q_\ell$ and a matching m in the original market with $\Delta_{i,m(i)} > 0$. Let u assign its holders to seats in the order specified above. The target prefers $r(i)$ to $u(i)$, so apply the same tracing argument

from p_i. If it stops immediately, the target and its reference arm block m. Otherwise it ends at some p_{Q_h} with $h < \ell$, whose assigned seat is the reference seat of the preceding player in the chain and thus belongs to S'. The corresponding blocking pair in the original market involves $a_{\underline{m}(Q_h)}$ and a distinct assigned arm in S. Intermediate steps may compare two seats of the same arm, but the assignment of holders to seats rules out such a blocking pair at the endpoint. Thus every triplet used in the cover has a strictly positive gap between original arms.

To apply this cover to $\Gamma_i(T)$, define $\mathcal{W}_j = \{k' \mid a_{k'} \in S,\ \mu_{j,k'} < \mu_{j,\underline{m}(j)}\}$. For every matching m with positive regret gap, the tracing argument gives

$$\Delta_{i,m(i)} \le \sum_{k':\Delta_{i,k'}>0} \Delta_{i,k'} \mathbf{1}_{\{m \in B_{i,\underline{m}(i),k'}\}} + \sum_{h<\ell} \sum_{k' \in \mathcal{W}_{Q_h}} \mathbf{1}_{\{m \in B_{Q_h,\underline{m}(Q_h),k'}\}}.$$

If the tracing stops at the target, the corresponding term in the first sum equals $\Delta_{i,m(i)}$. Otherwise, a term in the second sum equals one, which is at least $\Delta_{i,m(i)}$. All remaining terms are nonnegative. Grouping the blocking pairs by original arms counts each original triplet once, even if it corresponds to several seats.

Fix any feasible solution $(x_m)_{m \in \mathcal{M}_c}$ for $\Gamma_i(T)$ and discard its nonpositive objective terms. Multiply this inequality by x_m and sum over matchings with positive regret gaps. Interchanging the finite sums gives a weighted sum of triplet counts over matchings with positive regret gaps. Since $x_m \ge 0$, each count is at most $\sum_{m \in B_{j,k,k'}} x_m$. Applying the triplet constraints with the same weights, omitting empty triplet sets, and maximizing over feasible solutions gives

$$\Gamma_i(T) \le \sum_{k':\Delta_{i,k'}>0} \left(5\Delta_{i,k'} + \frac{6\log T}{\Delta_{i,k'}}\right) + \sum_{h<\ell} \sum_{k' \in \mathcal{W}_{Q_h}} \left(5 + \frac{6\log T}{(\mu_{Q_h,\underline{m}(Q_h)} - \mu_{Q_h,k'})^2}\right). \quad (58)$$

Let r_j be the rank of $a_{\underline{m}(j)}$ in player p_j's preferences over the original arms. Every arm preferred to it belongs to S, since otherwise an empty arm would block $\underline{m}$. Thus $r_j \le s$ and exactly $s - r_j$ arms in S are worse. List these arms as $a_{k'_{j,1}}, \dots, a_{k'_{j,s-r_j}}$ in decreasing reward order. The rank of $a_{k'_{j,h}}$ among all arms is at least $r_j + h$, so

$$\mu_{j,\underline{m}(j)} - \mu_{j,k'_{j,h}} \ge \sum_{v=r_j}^{r_j+h-1} (\mu_{j,\sigma_{j,v}} - \mu_{j,\sigma_{j,v+1}}) \ge h\Delta,$$

where the last inequality follows from Definition 1, since $r_j + h - 1 \le s - 1 \le D$. Every positive target gap is at least Δ, so the first sum in (58) contributes at most $(K-1)(5 + 6\log T/\Delta)$. For each possible endpoint p_j of the chain,

$$\sum_{k' \in \mathcal{W}_j} \frac{1}{(\mu_{j,\underline{m}(j)} - \mu_{j,k'})^2} \le \frac{1}{\Delta^2} \sum_{h=1}^{s-r_j} \frac{1}{h^2} \le \frac{\pi^2}{6\Delta^2}.$$

Also, $|\mathcal{W}_j| = s - r_j \le s - 1$. There are at most $N - 1$ players preceding the target in the right order, so (58) gives

$$\Gamma_i(T) \le (K-1)\left(5 + \frac{6\log T}{\Delta}\right) + 5(N-1)(s-1) + \frac{\pi^2(N-1)\log T}{\Delta^2}. \quad (59)$$

Since $s - 1 \le D \le 1/\Delta$, the right-hand side is $O(K\log T/\Delta + N\log T/\Delta^2)$. Together with (53), this gives the regret bound under the α-condition in Theorem 10.

K Experiments

We give the full experimental setup, computational details, and additional results for the comparison in Section 7.

K.1 Experimental Setup

We fix $N \in \{2, 3, 4, 5\}$ and $K \in \{5, 10\}$ and generate 20 markets for each pair (N, K). All players' and arms' preference rankings are sampled independently as uniform permutations. We use three reward settings, keeping the same preference rankings for each market.

In the first two settings, the mean rewards have uniform adjacent gaps. Using $\sigma_{j,r}$ from Definition 1, we set $\mu_{j,\sigma_{j,r}} = 0.5 + \eta\left(\frac{K+1}{2} - r\right)$, with $\eta = 0.1$ or $\eta = 0.05$.

In the third setting, we draw each adjacent gap $\eta_{j,r}$ independently and uniformly from $\{0.05, 0.1\}$ for $j \in [N]$ and $r \in [K-1]$. We set $\mu_{j,\sigma_{j,r}} = 0.05 + \sum_{s=r}^{K-1} \eta_{j,s}$. Thus $\eta_{j,r}$ is the mean reward gap between arms at ranks r and $r+1$, and all means lie in $[0.05, 0.95]$. The gaps remain fixed across runs of the same market.

Rewards are independent Bernoulli variables with these means. For each market and reward setting, we run UCB-GS for $T = 100{,}000$ rounds over 50 independent repetitions. We report the regret and bounds for p_1.

K.2 Computation

In each run, we record $\sum_{s=1}^{t}\left(\mu_{1,\underline{m}(1)} - \mu_{1,m_s(1)}\right)$ and average over the 50 runs to estimate $R_1(t)$.

We compute all three bounds from their definitions, retaining the constants and enumerating all one-to-one matchings. For each arm with a positive regret gap, we compute the minimum cover cost in $\Lambda_1(t)$ using a binary integer program. Each binary variable selects a triplet, and every matching in $\mathcal{M}_{1,\ell}$ must be covered. We solve these programs using SciPy's `milp` with the relative optimality gap set to zero. For $\Gamma_1(t)$ and $\Psi_1(t)$, we use `linprog` with the HiGHS solver. We check solver success, coverage by the selected triplets, and feasibility of the returned linear program solutions.

The programs for $\Gamma_1(t)$ and $\Psi_1(t)$ include the horizon constraint. In all figures and tables, we also cap $\Lambda_1(t)$ at $t \max_{k\in[K]}\left(\mu_{1,\underline{m}(1)} - \mu_{1,k}\right)^+$, where $x^+ = \max\{x, 0\}$. This cap is applied separately in each market before averaging. The $\Lambda_1(t)$ values reported in Section 7 use the same cap.

For plots averaged over markets, we average the regret estimates and each bound over the 20 markets. Percentage reductions are computed from the bounds averaged over the 20 markets. Shading shows one standard error across markets, computed as the sample standard deviation of the 20 market averages divided by $\sqrt{20}$.

K.3 Results

Averages over markets. Figures 2, 3, and 4 compare the three bounds with empirical regret. At $T = 100{,}000$, the mean $\Gamma_1(T)$ is about 3.6% to 30.4% lower than the mean capped $\Lambda_1(T)$. The mean $\Psi_1(T)$ is about 3.8% to 33.7% lower than the same baseline. Relative to the mean $\Gamma_1(T)$, observation count sharing gives a further reduction of about 0.3% to 6.2%. Table 2 reports the ratios of the mean capped $\Lambda_1(T)$ to the mean $\Gamma_1(T)$. Reducing the uniform gap from 0.1 to 0.05 increases all three mean bounds at $T = 100{,}000$, while triplet count sharing gives similar relative reductions in both settings.

Individual markets. For each $N \in \{2, 3, 4, 5\}$, we selected one market uniformly at random from the 20 generated markets with $K = 10$ and $\eta = 0.1$. Figure 5 compares the three bounds for each selected market with empirical regret averaged over 50 independent runs. Shading shows one standard error across runs, computed as the sample standard deviation of regret divided by $\sqrt{50}$.

Table 2: Ratio of the mean capped $\Lambda_1(T)$ to the mean $\Gamma_1(T)$ at $T = 100{,}000$. Both means are computed over 20 markets.

N	K	Uniform 0.1	Uniform 0.05	Gaps in $\{0.05, 0.1\}$
2	10	1.062	1.062	1.037
3	10	1.160	1.160	1.172
4	10	1.437	1.437	1.364
5	10	1.357	1.357	1.425
2	5	1.133	1.132	1.122
3	5	1.163	1.163	1.113
4	5	1.263	1.289	1.257
5	5	1.190	1.190	1.152

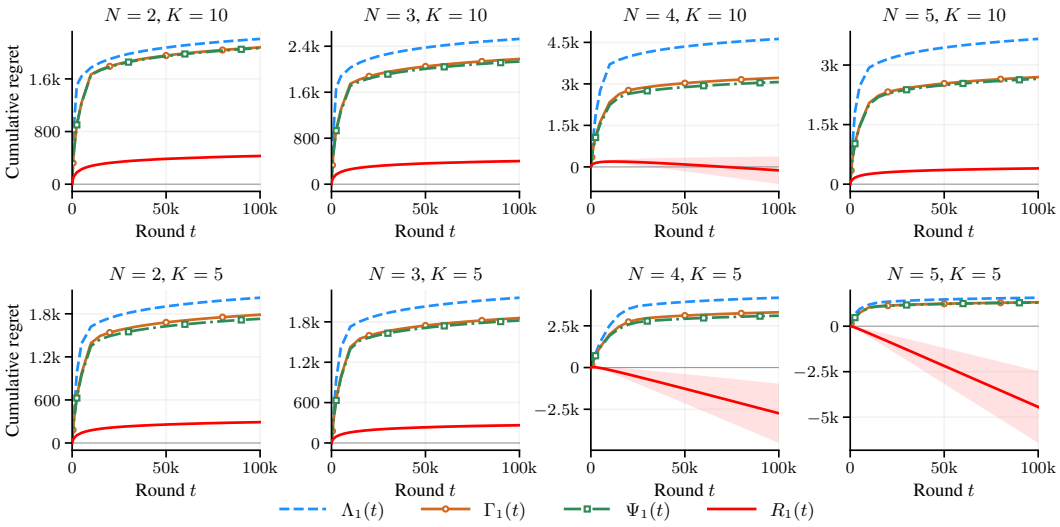

Figure 2: Regret bounds and empirical player-pessimal stable regret of UCB-GS with uniform adjacent mean reward gaps of 0.1.

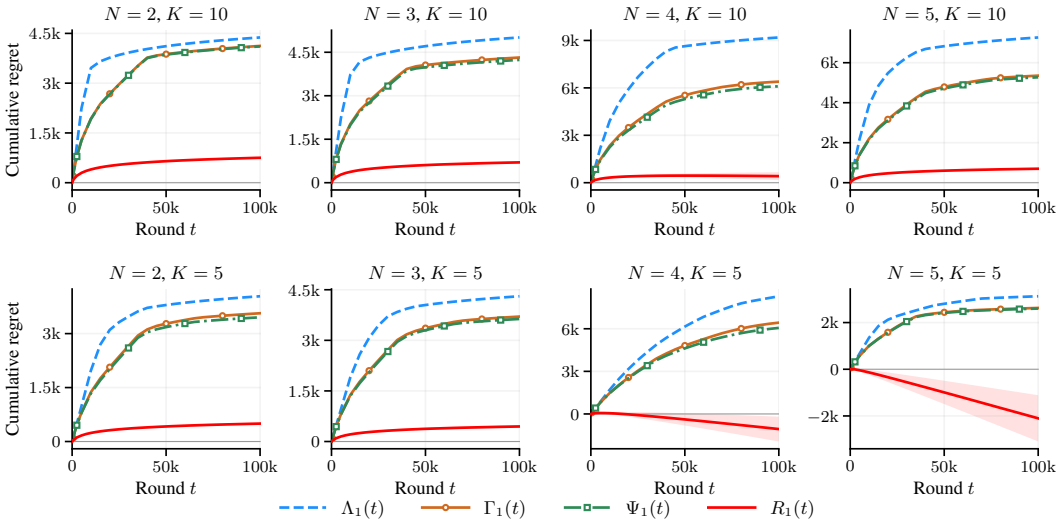

Figure 3: Regret bounds and empirical player-pessimal stable regret of UCB-GS with uniform adjacent mean reward gaps of 0.05.

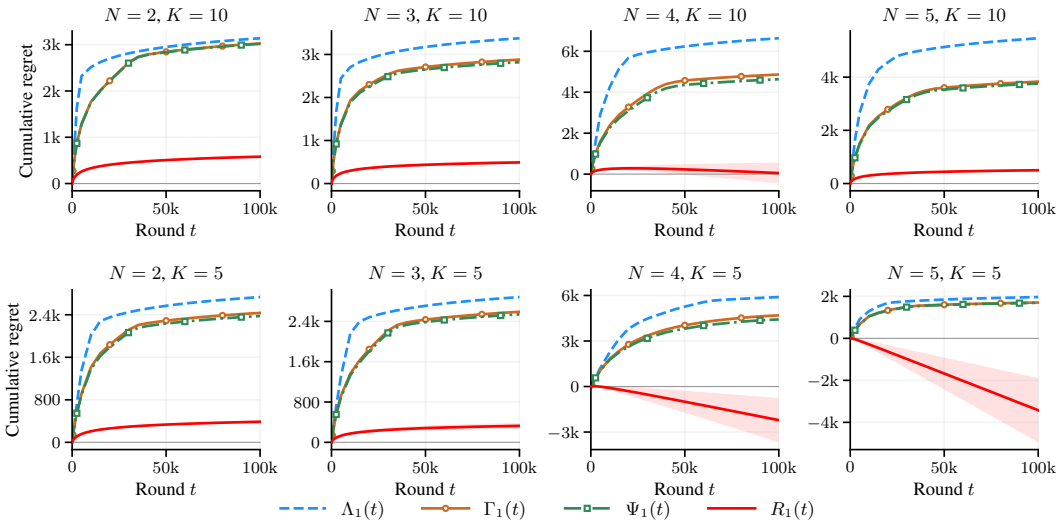

Figure 4: Regret bounds and empirical player-pessimal stable regret of UCB-GS with adjacent mean reward gaps drawn independently and uniformly from $\{0.05, 0.1\}$.

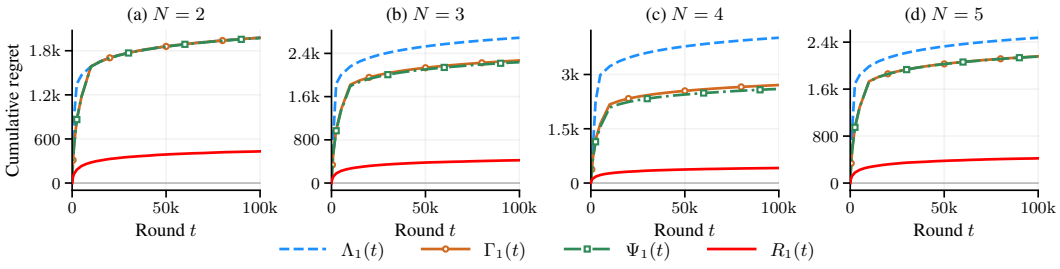

Figure 5: Regret bounds and empirical player-pessimal stable regret of UCB-GS on individual markets with $K = 10$ and uniform adjacent mean reward gaps of 0.1.