Private-Fair-Greedy: A Differentially Private Algorithm for Group Meritocratic Fairness in Contextual Linear Bandits

Esther An
Harvard University
ean@college.harvard.edu

Andrew Palacci
Harvard University
andrewpalacci@college.harvard.edu

Phevos Paschalidis
Harvard University
ppaschalidis@college.harvard.edu

May 8, 2025

Abstract

We develop an algorithm that simultaneously achieves privacy and fairness for the contextual linear bandit problem where candidates are associated with sensitive groups. In particular, we modify Grazzi et al.'s `Fair-Greedy` policy [GAF+22], which satisfies demographic parity at each round and obtains a high probability upper bound on fair pseudo-regret, to additionally achieve joint differential privacy (JDP). Our augmented policy incorporates differentially private steps for regression and rank estimation. We prove that the private and fair algorithm is (ε, δ)-JDP and evaluate its fairness and performance in experiments with U.S. Census data. We show empirically that the policy achieves sub-linear fair pseudo-regret, though the addition of noise leads to clear degradation of performance even under a longer horizon.

1 Introduction

In the classic *multi-armed bandit* (MAB) problem, a single agent aims to rapidly identify the optimal action (traditionally, arm) from a set of K alternatives by iteratively selecting among them and receiving noisy feedback. This well-studied framework for sequential decision-making under uncertainty has widespread applications across dynamic pricing, recommendation systems, and healthcare [Tho33; LS20; Sli+19; SB18; BC+12].

In many settings, underlying assumptions about the environment may change from round to round, affecting the associated reward. In the *contextual bandit* problem, the agent observes a context at the start of round t, with the reward dependent on both the context and the selected action. If no additional structure is assumed (i.e., the rewards of arms under different contexts are entirely independent), then the contextual bandit problem requires learning a bandit-per-context. In order to facilitate learning transfer between contexts, the *contextual linear bandit* problem uses the context to map each action, or arm, to a known feature vector. The reward associated with each action then depends on the same linear function of this feature.

Consider the following motivating example. An employer sequentially (during each successive recruitment cycle) hires a single applicant from an evolving pool of K candidates, each of whom can be featurized based on their application. Upon candidate selection, the employer receives a scalar reward based on the new hire's performance, which we assume depends linearly on their featurized application. Though the employer cannot know how well each candidate will perform a priori, they can learn expected performance by evaluating similar candidates who were hired previously.

Recently, analyses of machine learning methods have expanded beyond the traditional lens of performance and complexity optimization to incorporate additional desiderata. Fairness and privacy are two intuitively appealing goals. Both are immediately applicable to our motivating example: if candidates belong to sensitive demographic groups, an employer may want to utilize a "fair" algorithm that treats candidates similarly across groups. Furthermore, the behavior of the learning algorithm should not reveal sensitive information about past candidates when the algorithm is deployed.

1.1 Our Contributions

In this paper, we develop a fair and private algorithm for the contextual linear bandit problem. To do so, we utilize definitions from the well-established fields of algorithmic fairness and differential privacy (DP). Specifically, we present the `Private-Fair-Greedy` algorithm, which simultaneously pursues *group meritocratic fairness* (GMF), where the algorithm learns to pick the presented candidate with the highest relative rank among its sensitive group, and *joint differential privacy* (JDP), where actions at time $t' \neq t$ have limited dependence on the candidates presented at time t. Though prior work has considered the contextual linear bandit problem under GMF and JDP separately ([GAF+22] and [SS18], respectively), ours is the first approach to consider these desiderata jointly. We follow our presentation of the `Private-Fair-Greedy` algorithm with a theoretical proof of its privacy guarantees and empirical simulations that compare its performance under the GMF criteria to Grazzi et al.'s non-private `Fair-Greedy` algorithm [GAF+22] and the standard `OFUL` algorithm introduced by Abbasi-Yadkori et al. [APS11].

The rest of the paper is organized as follows. After a review of related work in Section 2, we present our problem setting, including mathematical definitions for GMF and JDP, in Section 3. Section 4 presents our learning algorithm, `Private-Fair-Greedy`, while Section 5 discusses its JDP guarantee. Finally, we include empirical simulations of algorithmic performance in Section 6 and offer concluding thoughts in Section 7.

Notation. We use k to index the arms of the bandit and t for the time step. We denote by $\langle \cdot, \cdot \rangle$ the Euclidean inner product and by $[N]$ the set $\{1, \ldots, N\}$ where $N \in \mathbb{N}$ is a natural number. We use $\{x_i\}_{i \in \mathcal{I}}$ as notation for the set $\{x_i | i \in \mathcal{I}\}$ and, similarly, $\{x_i\}_{i=1}^N$ for $\{x_i | i \in [N]\}$. Given $\mathcal{S} = \{x_i\}_{i \in \mathcal{I}}$, we also denote the set $\mathcal{S} \setminus x_i$ as $\mathcal{S}_{-i}$.

2 Related Work

Contextual Linear Bandits. The contextual linear bandit problem dates to the turn of the century [ABL03; Aue02]. Li et al. offer a foundational regression-based approach for learning in this setting, estimating a preference vector using the agent's past interactions and then selecting the action with the highest probability of being optimal [LCL+10]. As is common in the bandit literature, their probability calculation incorporates an optimism term based on their confidence in the regression estimate. The `OFUL` algorithm of [APS11] improves upon this approach by constructing tighter, elliptical confidence intervals around the regression estimate.

Algorithmic Fairness and Fairness for Multi-Armed Bandits. Algorithmic fairness, which has emerged as a field within machine learning, evaluates algorithmic decision-making based on various mathematical formulations of fairness. Individual fairness defines fairness as the mapping of similar individuals to similar outcomes in the context of classification tasks [Dwo06; DHP+11; LRD+17]. In contrast, group fairness depends on the categorization of individuals given sensitive attributes such as race, gender, or socioeconomic status [DHP+11; HPP+16; KLR+18]. Demographic parity is a measure of group fairness that calls for the probability of choosing a candidate from a given sensitive group to be the same across all groups [DHP+11]. In the contextual linear bandit setting, [GAF+22] introduces group meritocratic fairness (GMF) as a definition of group fairness based on the notion that candidates should be evaluated in terms of their relative performance compared to others from the same sensitive group. Specifically, a GMF policy is one that, at each round, selects the candidate with the highest relative performance, or relative rank, within its sensitive group.

Differential Privacy. Differential privacy (DP) is a mathematical guarantee of individual privacy that maintains the requirement that an adversary should not be able to identify whether any given individual's data in a dataset is changed arbitrarily [DN03; Dwo06]. Mathematically, an ε-DP algorithm is one in which the distribution of outputs remains close for inputs differing on only one individual; equivalently, the "privacy loss" for each individual is bounded by ε. In our paper, we utilize two major relaxations of pure DP: approximate (ε, δ)-DP characterizes mechanisms that satisfy ε-DP with probability $1 - \delta$, and ρ zero-concentrated DP (zCDP) allows for 2ρ sub-Gaussian (but potentially unbounded) privacy loss [DKM+06; BS16]. Finally, DP has also been studied under settings of continual observation and applied repeatedly in the bandit context [DNP+10; TD16; HH22; HGF+24; SS18]. Most notably, Shariff and Sheffet give a differentially private algorithm for contextual linear bandits [SS18].

Joint Examination of Fairness and DP. Past work on differential privacy and fairness has progressed largely independently. In supervised learning, however, recent theoretical results have highlighted an important conflict between the concepts of privacy and fairness. [CGK+19] and [Aga21] demonstrate that, under pure ε-DP and strong notions of group fairness, no algorithm can achieve non-trivial accuracy. Relaxed definitions of fairness and privacy, however, may be achieved jointly. [GA24] introduces a private and fair learning algorithm for binary classification. In the bandit space, [SGJ24] introduces a federated contextual bandit algorithm with formal DP guarantees and fairness constraints, though its notion of fairness focuses on exposure parity across actions rather than within-group merit. Moreover, its privacy guarantees are for information shared across agents, rather than over played actions and observed rewards.

3 Problem Formulation

Consider a contextual linear bandit problem with K arms. At each time step $t \in [T]$, the agent is presented with a context c_t that defines a decision set of K feature vectors

$$\mathcal{D}_t := \{\phi(c_t, k)\}_{k\in[K]} = \{X_{k,t}\}_{k\in[K]}$$

with $\phi(c_t, k) = X_{k,t} \in \mathbb{R}^d$ for some dimension d. Each feature vector is associated with reward $y_{k,t} = \langle X_{k,t}, \theta^* \rangle + \eta_{k,t}$, where θ^* is an unknown *preference vector* and $\eta_{k,t}$ is a sub-Gaussian and zero-meaned random scalar. The agent selects action $k_t \in [K]$ and observes the noisy reward $y_{k_t,t}$.

In the standard contextual linear bandit problem, the goal of the agent is to maximize the expected cumulative reward $\sum_{t=1}^{T} \langle X_{k_t,t}, \theta^* \rangle$ or, equivalently, minimize the expected cumulative regret

$$\sum_{t=1}^{T} \left(\max_{k^* \in K} \langle X_{k^*,t}, \theta^* \rangle - \langle X_{k_t,t}, \theta^* \rangle \right),$$

defined as the difference in reward between the chosen policy and the optimal, where the latter is assumed to know the preference vector θ^*. If feature vectors represent individuals who are affiliated with sensitive groups, however, standard regret minimization may be incompatible with traditional notions of group fairness: a regret-minimizing policy may always or overwhelmingly select candidates from the same sensitive group, treating members of other groups 'unfairly'. Indeed, it may not even be possible to compare individuals from different groups in an intuitively fair way. Therefore, we seek a learning algorithm that selects candidates who are most performant *as compared only to others from their own group.*

Formally, we assume that each arm $k \in [K]$ is associated with a fixed sensitive group with unknown distribution X_k. At each time step t, candidate $\phi(c_t, k) = X_{k,t}$ is drawn independently from X_k for every $k \in [K]$: the context c_t uniquely defines which element from each group is presented to the agent. Then, $\langle X_k, \theta^* \rangle$ is the true reward distribution for group k. We denote its cumulative distribution function as $\mathcal{F}_k$: $\mathcal{F}_k(y) = \mathbb{P}(\langle X_k, \theta^* \rangle \leq y)$ for all $y \in \mathbb{R}$. Our fairness definition is based on the *relative rank* of each candidate x, defined as $\mathcal{F}_k(\langle x, \theta^* \rangle)$. The relative rank of x is thus the probability that another element from its sensitive group gives lower reward. Note that higher relative rank is associated with more performant candidates.

Definition 1 (Group Meritocratic Fairness [Definition 3.1 of [GAF+22]])**.** *A policy $\{k_t\}_{t=1}^{T}$ is group meritocratic fair (GMF) if, for all $t \in [T]$ and $k \in [K]$, the relative rank of the chosen candidate $X_{k_t,t}$ is greater than that of all others in the decision set $\mathcal{D}_t$. That is, for all $t \in [T]$ and $k \in [K]$,*

$$\mathcal{F}_{k_t}(\langle X_{k_t,t}, \theta^* \rangle) \geq \mathcal{F}_k(\langle X_{k,t}, \theta^* \rangle).$$

Given that both the $\mathcal{F}_k$ distributions and the θ^* preference vector are unknown, this fairness notion is impossible to satisfy uniformly over time. Instead, the goal of the fair contextual linear bandit setting is to design algorithms that learn to be fair. In particular, we minimize the following fair pseudo-regret definition:

Definition 2 (Fair Pseudo-Regret [Definition 3.2 of [GAF+22]])**.** *Let $T \in \mathbb{N}$ be the time horizon, $\{k_t\}_{t=1}^{T}$ the chosen policy, and $\{k_t^*\}_{t=1}^{T}$ a GMF policy. Then, the cumulative fair pseudo-regret is*

$$R(T) := \sum_{t=1}^{T} \Big(\mathcal{F}_{k_t^*}(\langle X_{k_t^*,t}, \theta^* \rangle) - \mathcal{F}_{k_t}(\langle X_{k_t,t}, \theta^* \rangle) \Big).$$

Any policy with sub-linear fair pseudo-regret (with respect to T) learns to be GMF.

Remark 3.1 (GMF Policies Satisfy Demographic Parity)**.** *In addition to assuming that $X_{k,t}$ are identically and independently sampled from X_k, we also consider X_k independent from $X_{k'}$ when*

$k \neq k' \in [K]$. Then, for all $t \in [T]$, the relative ranks $\{\mathcal{F}_k(\langle X_{k,t}, \theta^* \rangle)\}_{k \in [K]}$ are identically and independently distributed as uniform within $[0, 1]$ by the probability integral transform. Every GMF policy therefore satisfies (history-agnostic) demographic parity: given a history $\mathcal{H}_t := \cup_{i=1}^{t}\{\mathcal{D}_i, y_i, k_t\}$, we have $\mathbb{P}(k_t^* = k) = \mathbb{P}(k_t^* = k|\mathcal{H}_{t-1}) = 1/K$ for any GMF policy $\{k_t^*\}_{k \in [K]}$. This key property has important implications for algorithmic design, which we discuss in detail in Remark 4.1.

Even a GMF policy, however, can violate the privacy of observed and selected individuals. In the context of our motivating example, adversaries can collaborate to learn candidate information and the true performance of past hires based on an employer's subsequent decisions. Standard differential privacy notions are incompatible with MAB algorithms because the action of any optimal policy at time t is highly dependent on the dataset $\mathcal{D}_t$. We therefore rely on the definition of joint differential privacy introduced by [KPR+15]. Roughly, a mechanism is jointly differentially private (JDP) if, at every time step t, the joint distribution of outputs for steps $t' \neq t$ is stable with respect to differences in the dataset at time t.

More formally, let $z_t = \{y_{k,t}\}_{k \in [K]}$ be the set of possible rewards across all arms associated with dataset $\mathcal{D}_t$. We say two complete data histories $\mathcal{T} = \{\mathcal{D}_t, z_t\}_{t \in [T]}$ and $\mathcal{T}' = \{\mathcal{D}'_t, z'_t\}_{t \in [T]}$ be t-neighbors if it holds that $(\mathcal{D}_{t'}, z_{t'}) = (\mathcal{D}'_{t'}, z'_{t'})$ for all $t' \neq t$ and $(\mathcal{D}_t, z_t)$ and $(\mathcal{D}'_t, z'_t)$ differ by up to a single tuple $(X_{k,t}, y_{k,t})$ for some specific $k \in [K]$. That is, $\mathcal{T}$ and $\mathcal{T}'$ are adjacent if they differ by only a single candidate across all $t \in [T]$ and $k \in [K]$, specifically at time t.

Definition 3 (Joint Differential Privacy [Definition 6 of [KPR+15]])**.** *A randomized algorithm A is (ε, δ)-jointly differentially private if, for any t, any t-neighboring histories $\mathcal{T}$ and $\mathcal{T}'$, and any set of $T-1$ actions $B \in [K]^{T-1}$, it holds that*

$$\mathbb{P}(A(\mathcal{T})_{-t} \in B) \leq e^{\varepsilon}\mathbb{P}(A(\mathcal{T}')_{-t} \in B) + \delta.$$

At first glance, it may not be immediately clear how to reconcile this definition with our formulation for the fair contextual linear bandit problem, as the MAB agent at time t has access only to $(\mathcal{D}_1, y_{k_1}, \ldots, \mathcal{D}_{t-1}, y_{k_{t-1}}, \mathcal{D}_t)$ and not the complete data history. Nevertheless, any MAB algorithm can be formulated as taking in the complete history, restricting itself to make decisions based on only the information that is observable, and finally outputting a set of actions $\{k_t\}_{t=1}^{T}$. Definition 3 thus guarantees that the joint distribution over actions for time steps $t' \neq t$ for any MAB algorithm satisfying JDP is stable with respect to differences in the candidate set at time t. In contrast to group meritocratic fairness (Definition 1), which is not immediately satisfiable, we impose JDP as a hard constraint for our proposed algorithm.

4 The Private-Fair-Greedy Policy

We present the `Private-Fair-Greedy` policy. At each time step, the algorithm utilizes past data to compute a private regression estimate of the agent's preference vector, which it uses to further privately estimate the relative rank of available candidates. It then acts greedily, selecting the candidate with highest relative rank at each time step. `Private-Fair-Greedy` thus differs from the `Fair-Greedy` algorithm of [GAF+22] in its use of differentially private techniques for regression and rank estimation rather than standard ridge regression and simple counting. We discuss these steps in greater detail in Sections 4.1 and 4.2. The pseudocode for `Private-Fair-Greedy` is given in Algorithm 1.

Algorithm 1 Private-Fair-Greedy

```
Input: Time horizon $T$, privacy parameters $\varepsilon$ and $\delta$, boundedness parameter $\tilde{L}$, hyperparameters $\alpha_\varepsilon$ and $\alpha_\delta$, and algorithms Private-Regression and Private-Rank-Estimation.
1: Define $\varepsilon_{\text{reg}} = \varepsilon \cdot \alpha_\varepsilon, \delta_{\text{reg}} = \delta \cdot \alpha_\delta, \varepsilon_{\text{rank}} = \varepsilon - \varepsilon_{\text{reg}},$ and $\delta_{\text{rank}} = \delta - \delta_{\text{reg}}$.
2: for all $t \in [T]$ do
3:     Receive $\mathcal{D}_t = \{X_{k,t}\}_{k\in[K]}$.
4:     Set $\tilde{t} = \lfloor (t-1)/2 \rfloor$, $X_{1:\tilde{t}} = (X_{k_1,1}, \dots, X_{k_{\tilde{t}},\tilde{t}})^\top$, $y_{1:\tilde{t}} = (y_{k_1,1}, \dots, y_{k_{\tilde{t}},\tilde{t}})$, and $\mathcal{D}_{\tilde{t}+1:t} = (\mathcal{D}_{\tilde{t}+1}, \dots, \mathcal{D}_t)$.
5:     Calculate
           $$\hat{\theta}_{\tilde{t}} \leftarrow \texttt{Private-Regression}(X_{1:\tilde{t}}, y_{1:\tilde{t}}, \varepsilon_{\text{reg}}, \delta_{\text{reg}}, T)$$
6:     For each $k \in [K]$, calculate
           $$\hat{F}_{k,t}(\langle X_{k,t}, \hat{\theta}_{\tilde{t}}\rangle) \leftarrow \texttt{Private-Rank-Estimation}(\mathcal{D}_{\tilde{t}+1:t}, \hat{\theta}_{\tilde{t}}, \varepsilon_{\text{rank}}, \delta_{\text{rank}}, T).$$
7:     Sample action
           $$k_t = \arg\max_{k\in[K]} \hat{F}_{k,t}(\langle X_{k,t}, \hat{\theta}_{\tilde{t}}\rangle).$$
8:     Observe noisy reward $y_{k_t,t} = \langle X_{k_t,t}, \theta^* \rangle + \eta_t$.
9: end for
```

We manage the total privacy budget for the private algorithm by dividing ε and δ between the regression and relative rank estimation tasks. Then, our private regression and private rank estimation subroutines each allocate their respective budgets across the T time steps, as discussed in Sections 4.1 and 4.2. To justify this two-step, independent division, we rely on a continual composition theorem that shows that interactive continual tasks share the privacy guarantees of non-interactive tasks [HSV25]. This is necessary because our regression and rank estimation steps are intimately interleaved. We discuss the joint differential privacy guarantee of our algorithm in greater detail in Section 5.

Remark 4.1 (Exploration in the Private-Fair-Greedy Algorithm)**.** *One key principle that our* `Private-Fair-Greedy` *algorithm inherits from the* `Fair-Greedy` *approach of [GAF+22] is its greedy selection of the action with the highest estimated relative rank (Step 7). Most bandit algorithms explicitly avoid greediness in order to appropriately balance exploration and exploitation. The LinUCB and OFUL algorithms developed for the standard contextual linear bandit problem, for instance, maintain an elliptical confidence set around their regression estimate $\hat{\theta}$ that is wider in directions of greatest uncertainty [LCL+10; APS11]. By optimizing over all vectors within this set, the algorithms thus encourage actions in under-explored directions to prevent rapid convergence to a locally optimal but globally sub-optimal solution. As discussed in Remark 3.1, however, every GMF policy satisfies demographic parity: it selects an arm $k \in [K]$ with equal probability in each round t (unconditioned on $\mathcal{D}_t$). There is thus no tension between exploration and exploitation. Therefore, unlike standard bandit architectures, the fair contextual linear bandit problem offers exploration for free, since an optimally fair algorithm is also maximally exploratory.*

4.1 Continually Private Regression

In this section, we detail our technique for privately estimating the preference vector θ^* based on the feature vectors $X_{1:\tilde{t}}$ and rewards $y_{1:\tilde{t}}$ (Step 5 of Algorithm 1). Note that a non-private closed form solution for the associated linear ridge regression problem is

$$\hat{\theta}_{\tilde{t}}^{(\mathrm{np})} = (X_{1:\tilde{t}}^\top X_{1:\tilde{t}} + \lambda \mathbb{I}_d)^{-1} X_{1:\tilde{t}}^\top y_{1:\tilde{t}},$$

where λ is a regularization parameter. Intuitively, a private version of the regression must therefore add appropriate noise to both $X_{1:\tilde{t}}^\top X_{1:\tilde{t}}$ and $X_{1:\tilde{t}}^\top y_{1:\tilde{t}}$. Then, the resulting regression estimate will also be private by immunity to postprocessing.

We define $A := \begin{bmatrix} X_{1:T} & y_{1:T} \end{bmatrix} \in \mathbb{R}^{T\times(d+1)}$ with $A_{\tilde{t}}$ holding the top $\tilde{t}$ rows of A. Let $M_{\tilde{t}} = A_{\tilde{t}}^\top A_{\tilde{t}}$. Then, $X_{1:\tilde{t}}^\top X_{1:\tilde{t}}$ is the top left $d \times d$ sub-matrix of $M_{\tilde{t}}$, and $X_{1:\tilde{t}}^\top y_{1:\tilde{t}}$ makes up the first d entries of its last column. To perform a private regression, we thus need to maintain a noisy estimate of only $M_{\tilde{t}}$. To do so, we note that $M_{\tilde{t}}$ is an incrementally updated sum of the data at individual time steps since

$$M_{\tilde{t}+1} = M_{\tilde{t}} + \begin{bmatrix} X_{k_{\tilde{t}},\tilde{t}}^\top & y_{\tilde{t}} \end{bmatrix}^\top \begin{bmatrix} X_{k_{\tilde{t}},\tilde{t}}^\top & y_{\tilde{t}} \end{bmatrix}.$$

We can therefore use techniques from the framework of continual differential privacy to minimize noise addition. Specifically, we use the tree-based mechanism outlined in [CSS11] and [DNP+10], with the i^{th} leaf of the tree storing $\begin{bmatrix} X_{k_i,i}^\top & y_i \end{bmatrix}^\top \begin{bmatrix} X_{k_i,i}^\top & y_i \end{bmatrix}$. Each node maintains a noisy sum of the data stored in its subtree (i.e., a privacy-preserving intermediate sum $M_{i:j}$ for some i and j). Since any cumulative sum can be obtained by combining at most $m = 1 + \lceil \log(T/2) \rceil$ intermediate values (recall, the regression at time t uses the first $\tilde{t} \le t/2$ time steps), we are thus able to achieve $O(\varepsilon_{\text{reg}}/\sqrt{\log T})$ noise for each $M_{\tilde{t}}$ rather than the $O(\varepsilon_{\text{reg}}/\sqrt{T})$ factor offered by standard (non-continual) composition theorems.

Specifically, it suffices to preserve $(\varepsilon_{\text{reg}}/\sqrt{8m\ln(2/\delta_{\text{reg}})}, \delta_{\text{reg}}/2m)$-differential privacy at each node. We therefore add Gaussian noise with variance

$$\sigma_{\text{noise}}^2 := 16m\tilde{L}^4 \ln(4/\delta_{\text{reg}})^2/\varepsilon_{\text{reg}}^2$$

to each coordinate of the stored matrix, where $\tilde{L}$ is an upper bound on the norm of any $\begin{bmatrix} X_{k_{\tilde{t}},\tilde{t}} & y_{\tilde{t}} \end{bmatrix}$. That is, we sample $Z' \in \mathbb{R}^{(d+1)\times(d+1)}$ with $Z'_{i,j} \sim \mathcal{N}(0, \sigma_{\text{noise}}^2)$ independently and add $Z = (Z' + Z'^\top)/\sqrt{2}$ to each intermediate sum. To ensure that the resulting $M_{i:j}$ matrix remains positive semi-definite with high probability, we also shift by $2\Gamma\mathbb{I}_{d+1}$ with

$$\Gamma := \sigma_{\text{noise}}\sqrt{2m}(4\sqrt{d} + 2\ln(2T)).$$

Detailed pseudocode for the tree-based continual DP solution can be found in Algorithm 2 of [CSS11] and more information on our Gaussian noise derivation can be found in Section 4.2 of [SS18]. We thus formalize the DP guarantee of this continually private regression technique in Lemma 4.1, which follows from [SS18] and post-processing:

Lemma 4.1 (Private Regression is Private). *The sequence* $\{\theta_{\tilde{t}}\}_{\tilde{t}=1}^{\tilde{t}_{\max}}$ *where* $\tilde{t}_{\max} = \lfloor (T-1)/2 \rfloor$ *is* $(\varepsilon_{\text{reg}}, \delta_{\text{reg}})$*-DP with respect to the complete data history* $\mathcal{T} = \{(\mathcal{D}_t, z_t)\}_{t=1}^T$.

4.2 Private Relative Rank Estimation

We outline the `Private-Rank-Estimation` step of Algorithm 1, which produces a private estimation of relative rank for each arm at every round (Step 6). We observe that the non-private method of estimating relative rank presented by [GAF+22] simply counts the number of past candidates from the same sensitive group who have lower reward than the given candidate, with rewards estimated using the regression estimate $\hat{\theta}_{\tilde{t}}$. To privatize the relative rank of each candidate, it thus suffices to add appropriate Gaussian noise to the counting query.

More precisely, for each arm k, the non-private relative rank estimate at time t is:

$$\hat{F}_{k,t}^{(\mathrm{np})}(\langle X_{k,t}, \hat{\theta}_{\tilde{t}}\rangle) = \frac{1}{t-1-\tilde{t}} \sum_{s=\tilde{t}+1}^{t-1} \mathbb{1}\{\langle X_{k,s}, \hat{\theta}_{\tilde{t}}\rangle \leq \langle X_{k,t}, \hat{\theta}_{\tilde{t}}\rangle\}.$$

This is a simple count query of 1's divided by the known partial history length $N_t = (t-1) - \tilde{t}$, so we use the Gaussian mechanism for zero-concentrated differential privacy (zCDP) to privatize the mean estimation at each time step [BS16]. Specifically, we define

$$\sigma_t = \sqrt{\frac{T}{2N_t^2 \rho_{\mathrm{rank}}}}$$

where

$$\rho_{\mathrm{rank}} = \left(\sqrt{\log(1/\delta_{\mathrm{rank}}) + \varepsilon_{\mathrm{rank}}} - \sqrt{\log(1/\delta_{\mathrm{rank}})}\right)^2,$$

and release the private rank estimate

$$\hat{F}_{k,t}(\langle X_{k,t}, \hat{\theta}_{\tilde{t}}\rangle) = \hat{F}_{k,t}^{(\mathrm{np})}(\langle X_{k,t}, \hat{\theta}_{\tilde{t}}\rangle) + \mathcal{N}(0, \sigma_t^2) \tag{1}$$

for each $k \in [K]$.

Lemma 4.2 (Private Rank Estimation is Jointly Differentially Private)**.** *The sequence* $\{\{\hat{F}_{k,t}(\langle X_{k,t}, \hat{\theta}_{\tilde{t}}\rangle)\}_{k\in[K]}\}_{t\in[T]}$ *of rank estimates is* $(\varepsilon_{\mathrm{rank}}, \delta_{\mathrm{rank}})$*-JDP with respect to the complete data history* $\mathcal{T} = \{(\mathcal{D}_t, z_t)\}_{t=1}^T$.

Proof. By our definition of adjacency (see Definition 3), neighboring data histories differ only in a single tuple $(X_{k,t}, y_{k,t})$ across all $k \in [K]$ and $t \in [T]$. By parallel composition, as defined in [McS09], it therefore suffices to show that $\{\hat{F}_{k,t}(\langle X_{k,t}, \hat{\theta}_{\tilde{t}}\rangle)\}_{t\in[T]}$ is $(\varepsilon_{\mathrm{rank}}, \delta_{\mathrm{rank}})$-JDP for each $k \in [K]$ individually.

Let $k \in [K]$ be fixed. Given some $t \in [T]$, our private rank estimation outputs $\hat{F}_{k,t}(\langle X_{k,t}, \hat{\theta}_{\tilde{t}}\rangle)$ as in (1). Since indicators are in $\{0,1\}$, the sensitivity of $\hat{F}_{k,t}(\langle X_{k,t}, \hat{\theta}_{\tilde{t}}\rangle)$ is $\Delta q_t = 1/N_t$. Therefore, by Proposition 1.6 of [BS16], $\hat{F}_{k,t}(\langle X_{k,t}, \hat{\theta}_{\tilde{t}}\rangle)$ is ρ_0-zCDP for

$$\rho_0 = \frac{1}{2\sigma_t^2 \cdot N_t^2} = \frac{\rho_{\mathrm{rank}}}{T}.$$

By the composition properties of zCDP, then, the sequence $\{\hat{F}_{k,t}(\langle X_{k,t}, \hat{\theta}_{\tilde{t}}\rangle)\}_{t\in[T]}$ is ρ_{rank}-zCDP (Lemma 1.7 of [BS16]).

We note now that a ρ_{rank}-zCDP mechanism is $(\rho_{\text{rank}} + 2\sqrt{\rho_{\text{rank}} \log(1/\delta)}, \delta)$-DP for any $\delta > 0$ by Proposition 1.3 of [BS16]. Let $\upsilon = \log(1/\delta_{\text{rank}})$. Then, $\{\hat{F}_{k,t}(\langle X_{k,t}, \hat{\theta}_{\tilde{t}}\rangle)\}_{t\in[T]}$ is $(\varepsilon_{\text{rank}}, \delta_{\text{rank}})$-JDP for

$$\begin{aligned}
\varepsilon_{\text{rank}} &= \rho_{\text{rank}} + 2\sqrt{\rho_{\text{rank}} \cdot \upsilon} \\
&= \left(\sqrt{\upsilon + \varepsilon_{\text{rank}}} - \sqrt{\upsilon}\right)^2 + 2\sqrt{\left(\sqrt{\upsilon + \varepsilon_{\text{rank}}} - \sqrt{\upsilon}\right)^2 \cdot \upsilon} \\
&= \upsilon + \varepsilon_{\text{rank}} - 2\sqrt{\upsilon}\sqrt{\upsilon + \varepsilon_{\text{rank}}} + \upsilon + 2\sqrt{\upsilon}\left(\sqrt{\upsilon + \varepsilon_{\text{rank}}} - \sqrt{\upsilon}\right) \\
&= \varepsilon_{\text{rank}} + 2\upsilon - 2\sqrt{\upsilon}\sqrt{\upsilon + \varepsilon_{\text{rank}}} + 2\sqrt{\upsilon}\sqrt{\upsilon + \varepsilon_{\text{rank}}} - 2\upsilon \\
\varepsilon_{\text{rank}} &= \varepsilon_{\text{rank}}.
\end{aligned}$$

□

5 Joint Differential Privacy Proof

We are now ready to present our main theoretical result: `Private-Fair-Greedy` satisfies JDP.

Theorem 5.1 (`Private-Fair-Greedy` is Jointly Differentially Private)**.** *Given a contextual linear bandit problem with time horizon $T \in \mathbb{N}$ and feature vectors and rewards bounded such that $\|X_{k,t}\| + \|y_t\| \leq \tilde{L}$ for all $k \in [K]$ and $t \in [T]$, the `Private-Fair-Greedy` algorithm satisfies (ε, δ)-joint differential privacy.*

The proof follows mainly by noting that DP guarantees are immune to post-processing and by application of results in DP composition [DRV10; HSV25].

Proof Sketch. Elementary composition techniques allow us to consider the privacy loss associated with each time step t separately before combining them together to form a final bound. Using unions to represent composition, elementary arguments thus allow for the following deconstruction:

$$\text{Loss} = \bigcup_{t=1}^{T} \text{Loss}(t) = \bigcup_{t=1}^{T} \Big(\text{Loss}_{\text{reg}}(t) \cup \text{Loss}_{\text{rank}}(t)\Big).$$

Yet our analysis of privacy in Sections 4.1 and 4.2 are based on the composition of regression and rank estimation first over rounds and then between the two components. That is, ideally, we hope for composition of the form

$$\text{Loss} = \Big(\bigcup_{t=1}^{T} \text{Loss}_{\text{reg}}(t)\Big) \cup \Big(\bigcup_{t=1}^{T} \text{Loss}_{\text{rank}}(t)\Big),$$

in part because it allows for black box reduction to the privacy guarantees of [SS18] for our shared regression technique, and because it dramatically simplifies our argument. That this type of composition argument is indeed possible is a non-trivial observation: interactions between the private regression and private rank estimation steps of the algorithm (as the latter is based on the output of the former) could make the privacy guarantees of each inseparably connected. Nevertheless, Theorem 1.2 in [HSV25] guarantees exactly our desired result: the concurrent composition of interactive continual mechanisms is (ε, δ)-DP for the same (ε, δ) that hold for the composition of non-interactive mechanisms. Thus, we can first consider the cumulative privacy loss for private regression across all time steps (Lemma 4.1) before separately calculating the cumulative privacy loss for rank estimation (Lemma 4.2).

The remainder of the proof follows from basic composition and post-processing. We note that, at time t, the algorithm depends only on the regression estimate $\hat{\theta}_{\tilde{t}}$, the rank estimates $\{\hat{F}_{k,t}(\langle X_{k,t}, \hat{\theta}_{\tilde{t}}\rangle)\}_{t\in[T]}$, and the dataset $\mathcal{D}_t$ (see Step 7). By Lemma 4.1, the first variable is (ε, δ)-DP across all T, and, by Lemma 4.2, the second variable is (ε, δ)-JDP across all T. The final component depends only on the data at time t. Hence, it follows that `Private-Fair-Greedy` is (ε, δ)-JDP. □

6 Empirical Results

We evaluate the performance of `Private-Fair-Greedy` on the U.S. Census Adult dataset, as provided by the `folktables` library and following the pipeline introduced by [GAF+22]. This setting models a sequential "hiring" scenario, where an agent must select one candidate from a pool of K individuals (arms) per round. We set $K = G$, where each arm corresponds to one of G racial groups derived from the `RAC1P` attribute in the dataset, with only $G = 4$ groups represented given more than 50,000 samples retained. Context vectors are constructed from demographic features, including age, class of worker, education level, marital status, place of birth, relationship status, occupation, hours worked per week, sex, and race. In this "hiring" setting, the true reward for selecting a candidate is based on their potential income, which is modeled as a linear function of their context vector. Thus, the ground-truth coefficients for this linear function are learned via regression on a hold-out training set. This dataset and reward structure are identical to those used in the original [GAF+22] implementation, which facilitates a direct comparison between our private fair policy and the existing non-private fair policy.[1]

6.1 Setup & Baselines

We compare our algorithm against two baselines: the non-private `Fair-Greedy` algorithm of [GAF+22] and the standard `OFUL` algorithm [APS11]. `Fair-Greedy` explicitly optimizes for group meritocratic fairness. In contrast, `OFUL` maximizes expected reward (thus minimizing standard pseudo-regret) subject to confidence bounds.

All algorithms are evaluated over $T = 50{,}000$ rounds, repeated across 10 trials with distinct random seeds for noise generation. `Private-Fair-Greedy` is run with total privacy parameters $\varepsilon = 15$ and $\delta = 0.1$, as well as privacy-budgeting hyperparameters $\alpha_\varepsilon = \alpha_\delta = 0.9$, which refer to the proportion of the privacy budget allocated to the private regression step of the algorithm. We remark that ε and δ were chosen as empirically low values that still demonstrate (within a reasonable T) the learning of `Private-Fair-Greedy`. Furthermore, $\alpha_\varepsilon = \alpha_\delta$ were also chosen empirically, though more granular simulations could be run to optimize these hyperparameters given (ε, δ). Select examples of these empirical comparisons can be found in Figures A.1, A.2, A.3, A.4. Finally, we calculate the $\tilde{L}$ parameter from the dataset for continually private regression — we acknowledge that this step was not considered explicitly in our algorithm or privacy analysis, noting instead that this calculation exists to represent an assumption that our dataset is defined given known bounds.

6.2 Fairness Across Groups

Figure 1 presents the percentage of times an arm belonging to each group ($G1$ through $G4$) was selected across $T = 50{,}000$ rounds for each algorithm. The dotted line at 25% represents the expected

[1] Our repository (https://github.com/andrewp2303/DP-GMFBandits) extends [GAF+22]'s public codebase.

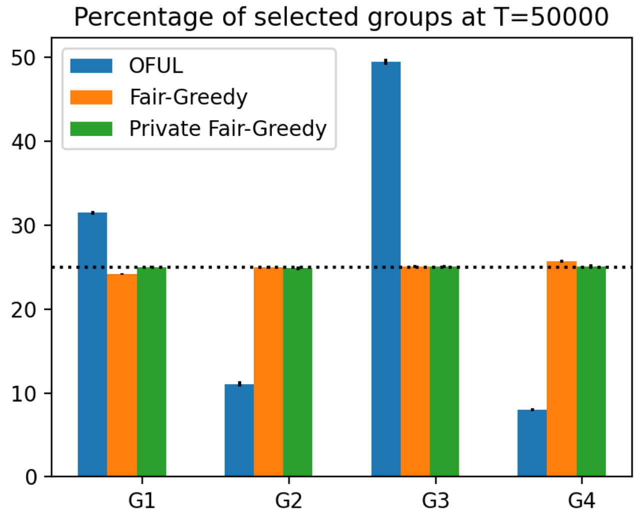

Figure 1: Percentage of times an arm from a given group ($G1$–$G4$) was selected by each algorithm. Dotted line indicates 25%, corresponding to $100\%/K$ where $K = 4$ is the number of arms.

selection rate per group under perfect demographic parity, given that one arm is selected from among a pool of $K = 4$ in each round. As anticipated, `Fair-Greedy` and `Private-Fair-Greedy` closely approximate this parity across groups. In contrast, but as expected, `OFUL`, which does not incorporate fairness constraints, demonstrates significant bias by frequently selecting arms from one group ($G3$) while neglecting others ($G2$, $G4$).

6.3 Fair Pseudo-Regret

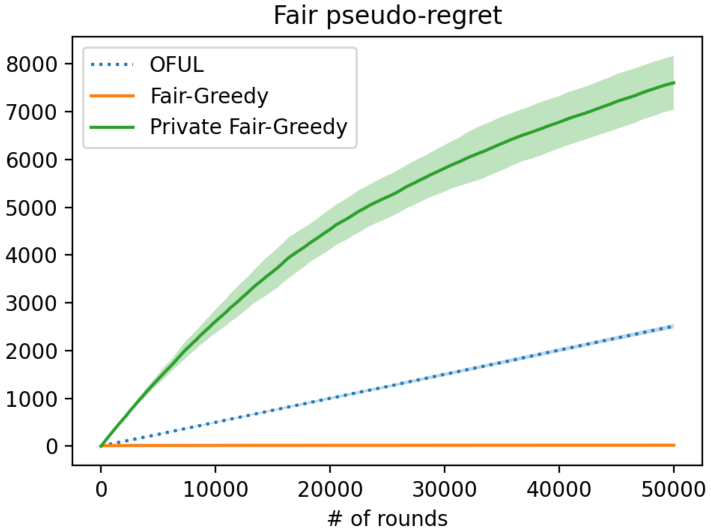

Figure 2: Cumulative fair pseudo-regret for each algorithm over $T = 50{,}000$ rounds.

We next measure deviation from group meritocratic fairness using cumulative fair pseudo-regret

(Definition 2). Figure 2 plots this measurement for each algorithm against the number of rounds T.

These results demonstrate the distinct behavior of each of the three algorithms. The non-private `Fair-Greedy` policy accrues near-zero cumulative fair pseudo-regret over the horizon, indicating that its selections closely approximate those of an optimal GMF policy. The standard `OFUL` policy shows approximately linear growth in fair pseudo-regret, which makes intuitive sense: without a notion of fairness incorporated into the policy's reward measurements, we should expect consistent violations of GMF over time. Finally, the `Private-Fair-Greedy` policy (like the non-private version) exhibits sublinear growth in cumulative fair pseudo-regret — however, within the evaluated horizon of $T = 50{,}000$ rounds, the fair pseudo-regret is *much* higher than that of `OFUL`, suggesting that the noise we introduce causes serious degradation within this horizon, even with the privacy loss parameters $\varepsilon = 15, \delta = 0.1$.

6.4 Limitations

While our empirical results demonstrate the behavior of `Private-Fair-Greedy` in a semi-realistic "hiring" scenario, it is important to acknowledge several limitations of the current evaluation. First, the scale of the experimental setup, given a relatively small number of arms and corresponding sensitive groups ($K = G = 4$), may not accurately reflect the challenges of real-world applications that involve significantly more candidates per round (e.g. in the context of a different sensitive attribute) or consider finer-grained and intersectional group distinctions. Generalizing the observed results to these more complex settings appears feasible given the current dataset and problem setting but is beyond the immediate scope of what we aimed to accomplish with this work.

Additionally, our evaluation is conducted over a fixed horizon of $T = 50{,}000$ rounds. This bound is primarily due to computational constraints, particularly those associated with the relative rank estimation step for `Fair-Greedy` and `Private-Fair-Greedy`. Addressing this bottleneck and running the algorithm over a longer horizon would be helpful for further understanding `Private-Fair-Greedy`'s asymptotic behavior and allow for greater exploration.

7 Conclusion & Future Work

In this project, we extend the implementation of the `Fair-Greedy` algorithm for the contextual linear bandit problem presented by [GAF+22] to pursue not only group meritocratic fairness but also joint differential privacy. We motivate the implementation of a differential privacy guarantee for the algorithm in the context of its real-world use case and show that our `Private-Fair-Greedy` algorithm satisfies (ε, δ)-JDP. We also provide an empirical evaluation of the fairness and performance of the `Private-Fair-Greedy` algorithm, as well as comparisons to the original `Fair-Greedy` algorithm and `OFUL`.

A natural extension of our theoretical results is to provide regret analysis for Algorithm 1. This analysis is beyond the immediate scope of this course, but would nonetheless be valuable for theoretically interpreting the performance of the `Private-Fair-Greedy` policy. Moreover, we note that we provide one method for privatizing the `Fair-Greedy` algorithm of [GAF+22] to satisfy the definition of joint differential privacy through specific implementations of a continually private regression and private relative rank estimation. We recognize that there exist other methods for privatizing these components of the algorithm, and, moreover, implemented several; alternative implementations of private ridge regression and relative rank estimation remain in the repository history, with a theoretical analysis of these implementations' privacy guarantees left to future work. Furthermore, `Private-Fair-Greedy` involves the implementation of a method for continually private regression but not continually private rank estimation — we envision a potential extension of this algorithm to involve the latter as well.

Given the relevance of the contextual linear bandit setting to high-stakes fields where consequential decisions are made about individuals using their sensitive attributes, it remains highly important to examine how algorithms in this setting can provide guarantees of both privacy and fairness. We hope that continued work at this intersection will contribute to the development of algorithms that satisfy normative desiderata for safe and effective machine learning.

Contributions

Andrew: Problem Formulation, Private Relative Rank Estimation, Empirical Results, Appendix, all code implementation and experimentation (see Github repo), revisions.

Esther: Abstract, Related Work, Private Relative Rank Estimation, Regression, Conclusion & Future Work, Poster, zCDP Proof, Continually Private Regression code, revisions.

Phevos: Introduction, Problem Formulation, Private-Fair-Greedy Policy overview, Continually Private Regression, Joint DP Proof, Continually Private Regression code, revisions.

References

[GAF+22] Riccardo Grazzi, Arya Akhavan, John IF Falk, Leonardo Cella, and Massimiliano Pontil. "Group meritocratic fairness in linear contextual bandits". In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 24392–24404.

[Tho33] William R. Thompson. "On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples". In: *Biometrika* 25.3/4 (1933), pp. 285–294. ISSN: 00063444. URL: `http://www.jstor.org/stable/2332286` (visited on 09/23/2023).

[LS20] Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

[Sli+19] Aleksandrs Slivkins et al. "Introduction to multi-armed bandits". In: *Foundations and Trends® in Machine Learning* 12.1-2 (2019), pp. 1–286.

[SB18] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[BC+12] Sébastien Bubeck, Nicolo Cesa-Bianchi, et al. "Regret analysis of stochastic and nonstochastic multi-armed bandit problems". In: *Foundations and Trends® in Machine Learning* 5.1 (2012), pp. 1–122.

[SS18] Roshan Shariff and Or Sheffet. "Differentially private contextual linear bandits". In: *Advances in Neural Information Processing Systems* 31 (2018).

[APS11] Yasin Abbasi-yadkori, Dávid Pál, and Csaba Szepesvári. "Improved Algorithms for Linear Stochastic Bandits". In: *Advances in Neural Information Processing Systems*. Ed. by J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K.Q. Weinberger. Vol. 24. Curran Associates, Inc., 2011. URL: https://proceedings.neurips.cc/paper_files/paper/2011/file/e1d5be1c7f2f456670de3d53c7b54f4a-Paper.pdf.

[ABL03] Naoki Abe, Alan W Biermann, and Philip M Long. "Reinforcement learning with immediate rewards and linear hypotheses". In: *Algorithmica* 37 (2003), pp. 263–293.

[Aue02] Peter Auer. "Using confidence bounds for exploitation-exploration trade-offs". In: *Journal of Machine Learning Research* 3.Nov (2002), pp. 397–422.

[LCL+10] Lihong Li, Wei Chu, John Langford, and Robert E Schapire. "A contextual-bandit approach to personalized news article recommendation". In: *Proceedings of the 19th international conference on World wide web*. 2010, pp. 661–670.

[Dwo06] Cynthia Dwork. "Differential Privacy". In: *Automata, Languages and Programming*. Ed. by Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 1–12. ISBN: 978-3-540-35908-1.

[DHP+11] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Rich Zemel. *Fairness Through Awareness*. 2011. arXiv: 1104.3913 [cs.CC]. URL: https://arxiv.org/abs/1104.3913.

[LRD+17] Yang Liu, Goran Radanovic, Christos Dimitrakakis, Debmalya Mandal, and David C. Parkes. *Calibrated Fairness in Bandits*. 2017. arXiv: 1707.01875 [cs.LG]. URL: https://arxiv.org/abs/1707.01875.

[HPP+16] Moritz Hardt, Eric Price, Eric Price, and Nati Srebro. "Equality of Opportunity in Supervised Learning". In: *Advances in Neural Information Processing Systems*. Ed. by D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett. Vol. 29. Curran Associates, Inc., 2016. URL: https://proceedings.neurips.cc/paper_files/paper/2016/file/9d2682367c3935defcb1f9e247a97c0d-Paper.pdf.

[KLR+18] Matt J. Kusner, Joshua R. Loftus, Chris Russell, and Ricardo Silva. *Counterfactual Fairness*. 2018. arXiv: 1703.06856 [stat.ML]. URL: https://arxiv.org/abs/1703.06856.

[DN03] Irit Dinur and Kobbi Nissim. "Revealing information while preserving privacy". In: *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*. PODS '03. San Diego, California: Association for Computing Machinery, 2003, pp. 202–210. ISBN: 1581136706. DOI: 10.1145/773153.773173. URL: https://doi.org/10.1145/773153.773173.

[DKM+06] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. "Our data, ourselves: privacy via distributed noise generation". In: *Proceedings of the 24th Annual International Conference on The Theory and Applications of Cryptographic Techniques*. EUROCRYPT'06. St. Petersburg, Russia: Springer-Verlag, 2006, pp. 486–503. ISBN: 3540345469. DOI: 10.1007/11761679_29. URL: https://doi.org/10.1007/11761679_29.

[BS16] Mark Bun and Thomas Steinke. "Concentrated differential privacy: Simplifications, extensions, and lower bounds". In: *Theory of cryptography conference*. Springer. 2016, pp. 635–658.

[DNP+10] Cynthia Dwork, Moni Naor, Toniann Pitassi, and Guy N. Rothblum. “Differential privacy under continual observation”. In: *Proceedings of the Forty-Second ACM Symposium on Theory of Computing*. STOC ’10. Cambridge, Massachusetts, USA: Association for Computing Machinery, 2010, pp. 715–724. ISBN: 9781450300506. DOI: 10.1145/1806689.1806787. URL: https://doi.org/10.1145/1806689.1806787.

[TD16] Aristide Tossou and Christos Dimitrakakis. “Algorithms for differentially private multi-armed bandits”. In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 30. 1. 2016.

[HH22] Bingshan Hu and Nidhi Hegde. “Near-optimal Thompson sampling-based algorithms for differentially private stochastic bandits”. In: *Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence*. Ed. by James Cussens and Kun Zhang. Vol. 180. Proceedings of Machine Learning Research. PMLR, Jan. 2022, pp. 844–852. URL: https://proceedings.mlr.press/v180/hu22a.html.

[HGF+24] Osama Hanna, Antonious M Girgis, Christina Fragouli, and Suhas Diggavi. “Differentially private stochastic linear bandits:(almost) for free”. In: *IEEE Journal on Selected Areas in Information Theory* (2024).

[CGK+19] Rachel Cummings, Varun Gupta, Dhamma Kimpara, and Jamie Morgenstern. “On the Compatibility of Privacy and Fairness”. In: *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*. UMAP’19 Adjunct. Larnaca, Cyprus: Association for Computing Machinery, 2019, pp. 309–315. ISBN: 9781450367110. DOI: 10.1145/3314183.3323847. URL: https://doi.org/10.1145/3314183.3323847.

[Aga21] Sushant Agarwal. “Trade-Offs between Fairness and Privacy in Machine Learning”. eng. In: 2021.

[GA24] Hrad Ghoukasian and Shahab Asoodeh. “Differentially private fair binary classifications”. In: *2024 IEEE International Symposium on Information Theory (ISIT)*. IEEE. 2024, pp. 611–616.

[SGJ24] Sambhav Solanki, Sujit Gujar, and Shweta Jain. “Fairness and Privacy Guarantees in Federated Contextual Bandits”. In: *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems*. AAMAS ’24. Auckland, New Zealand: International Foundation for Autonomous Agents and Multiagent Systems, 2024, pp. 2471–2473. ISBN: 9798400704864.

[KPR+15] Michael Kearns, Mallesh M Pai, Ryan Rogers, Aaron Roth, and Jonathan Ullman. “Robust mediators in large games”. In: *arXiv preprint arXiv:1512.02698* (2015).

[HSV25] Monika Henzinger, Roodabeh Safavi, and Salil Vadhan. *Concurrent Composition for Differentially Private Continual Mechanisms*. 2025. arXiv: 2411.03299 [cs.DS]. URL: https://arxiv.org/abs/2411.03299.

[CSS11] T-H Hubert Chan, Elaine Shi, and Dawn Song. “Private and continual release of statistics”. In: *ACM Transactions on Information and System Security (TISSEC)* 14.3 (2011), pp. 1–24.

[McS09] Frank D. McSherry. “Privacy integrated queries: an extensible platform for privacy-preserving data analysis”. en. In: *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data*. Providence Rhode Island USA: ACM, June 2009, pp. 19–30. ISBN: 978-1-60558-551-2. DOI: 10.1145/1559845.1559850. URL: https://dl.acm.org/doi/10.1145/1559845.1559850.

[DRV10] Cynthia Dwork, Guy N Rothblum, and Salil Vadhan. “Boosting and Differential Privacy”. en. In: *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science* FOCS ’10 (2010), pp. 51–60.

A Appendix: Additional Experimental Results

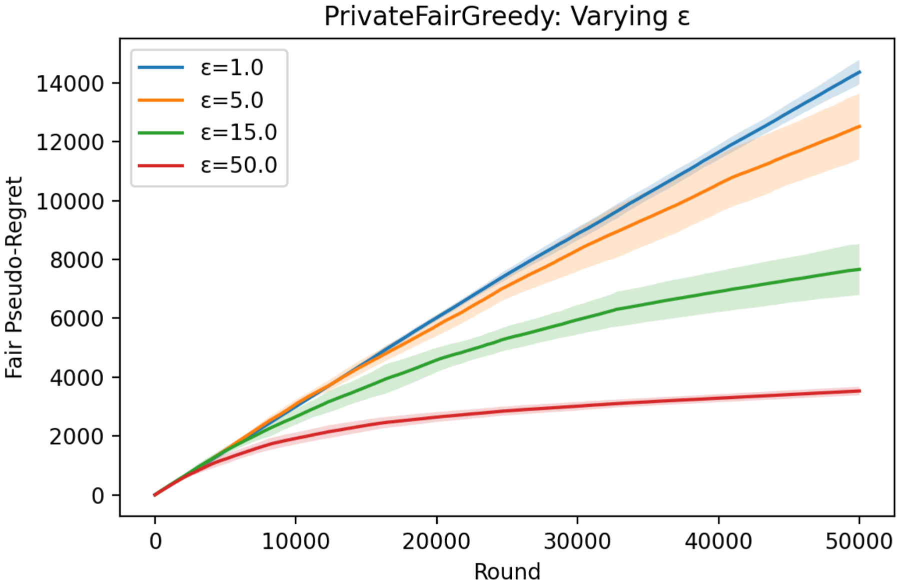

Figure A.1: Effect of $\varepsilon \in \{1, 5, 15, 50\}$ on fair pseudo-regret (with fixed $\delta = 0.1, \alpha_\varepsilon = 0.9, \alpha_\delta = 0.9$). As ε increases, regret decreases, confirming the expected privacy-utility tradeoff.

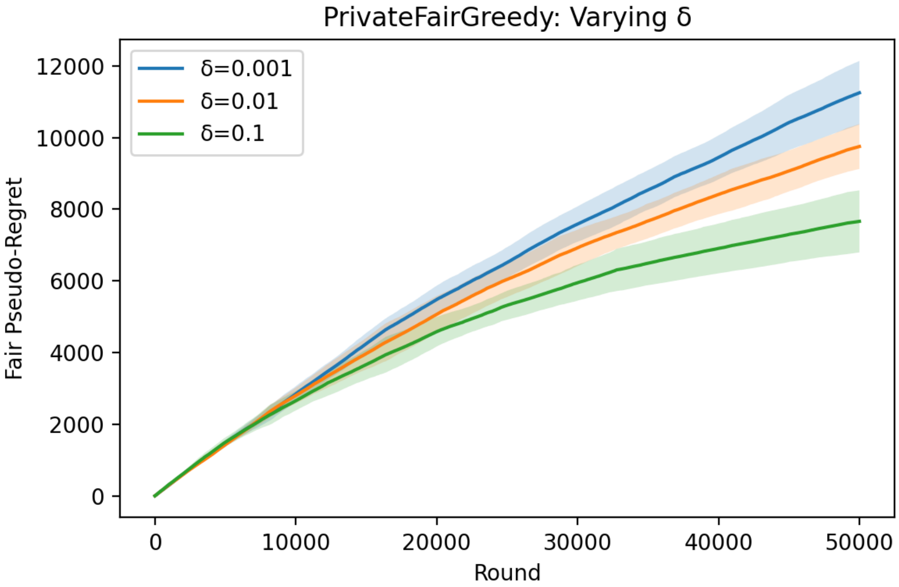

Figure A.2: Effect of $\delta \in \{0.001, 0.01, 0.1\}$ on fair pseudo-regret (with fixed $\varepsilon = 15, \alpha_\varepsilon = 0.9, \alpha_\delta = 0.9$). Larger δ's allow more slack for privacy violations, resulting (as expected) in lower fair pseudo-regret.

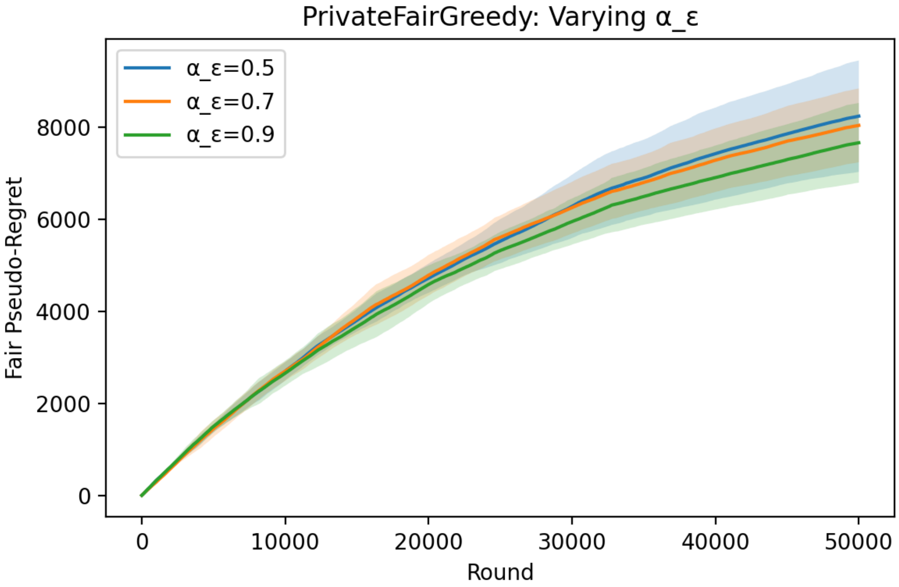

Figure A.3: Effect of $\alpha_\varepsilon \in \{0.5, 0.7, 0.9\}$ on fair pseudo-regret (with fixed $\varepsilon = 15, \delta = 0.1, \alpha_\delta = 0.9$). This controls how the privacy budget is split between regression and rank estimation, so we find that greater values (more budget for regression) yield slightly lower fair pseudo-regret.

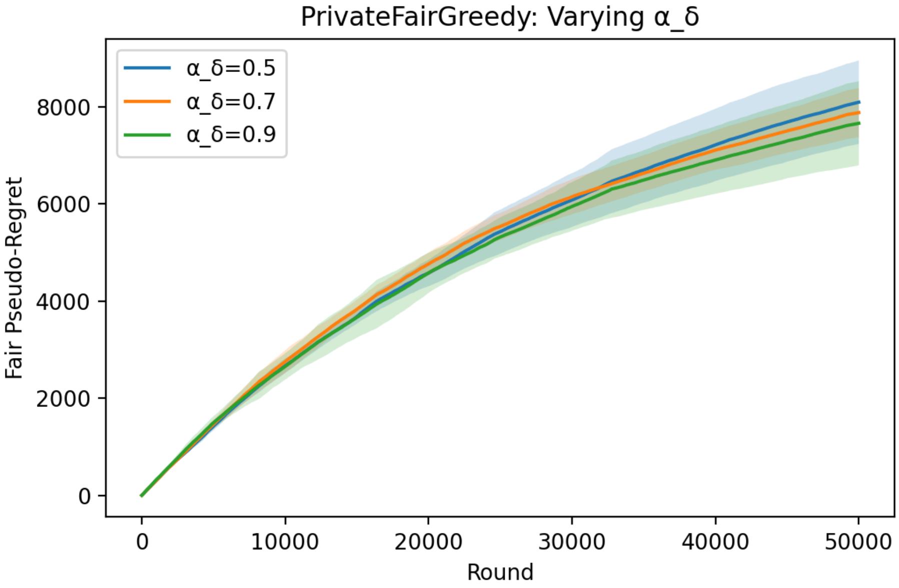

Figure A.4: Effect of $\alpha_\delta \in \{0.5, 0.7, 0.9\}$ on fair pseudo-regret (with fixed $\varepsilon = 15, \delta = 0.1, \alpha_\delta = 0.9$). As with α_ε, we again find that greater values yield lower fair pseudo-regret.