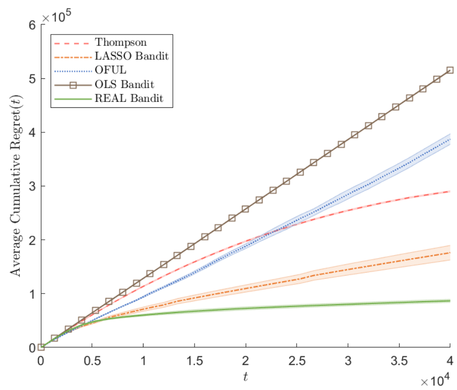

We thank all reviewers for their valuable suggestions on citing relevant literature and improving organization of the paper. Responses to other questions are presented below.

Simulations. Reviewer 1 found the paper more on the theoretical side and also asked about practicality of our theory-driven algorithm, and Reviewer 2 requested a numerical study. We compared REAL Bandit with 4 other algorithms: OLS-Bandit of [15][1] , LASSO-Bandit of [6], OFUL of [2] which is UCB based, and Thompson sampling (from [36]). We generated matrix $\mathbf{B}^\star$ as $\mathbf{U}\mathbf{V}^\top$ where $\mathbf{U} \in \mathbb{R}^{201\times 3}$ and $\mathbf{V} \in \mathbb{R}^{200\times 3}$ with iid $\mathcal{N}(0,1)$ entries. Noise variance is 1 and features are iid $\mathcal{N}(\mathbf{0}, \mathbf{I}_d)$. We gave Thompson sampling the true prior mean and variance of arm parameters, and true noise variance. We generated 10 data sets, ran all algorithms, and present their average cumulative regret (with 1 SE error bars) for a time horizon of length $T = 40,000$ in the above figure. We are grateful to the reviewers for this suggestion since the simulations back our theoretical results.

Reviewer 1. *Practicality of the gap assumption:* unlike [2] or [36], we do not assume a deterministic gap exists (this would actually not hold since covariates can be very close to the decision boundaries). Our Assumption 3 of §A (adapted from [6,15]) only requires that a subset of arms are optimal with positive probability, and the remaining arms are sub-optimal with positive probability. It can be shown this assumption holds for all standard distributions for the covariates. *Optimality of factor r^2 in the regret:* we thank the reviewer for this, since it led us to a careful investigation of the bounds which made us realize the bounds can actually be tightened to replace r^2 with r, matching the bounds one sees in matrix completion literature.

Reviewer 2. *Theoretical contributions beyond [6,15]:* we highlight these major advances: 1) We provide a stronger characteristic of "all-sampling" observations which helps obtaining tail-bound inequalities for the all-sampling estimator (see function G in Assumption 12 of §C). The analysis in [6,15] are not sufficient for our low-rank estimator bound. 2) We allow for a randomized forced-sampling rule which is more flexible in practice than the deterministic sampling rules introduced in [6, 15]. 3) We proposed a simple method for identifying the optimal arms which is required for the analysis of estimators to work. *Explanation of Lemmas 5-6 and their relationship:* Lemmas 6-7 verify the assumptions of row-enhancement bounds while Lemma 5 verifies those of the trace-regression estimator. The former is concerned about the samples for *each individual* arm, whereas the latter does not care about individual arms and only demands iid samples among $\mathbf{X}_i$'s (which include samples from multiple arms). However, they are very similar in nature.

Reviewer 3. *Significance of the REAL-estimator:* low-rank tail bounds are known for $\|\widehat{\mathbf{B}} - \mathbf{B}^\star\|_F$, but we need row-wise bounds $\|\widehat{B}_\kappa - B^\star_\kappa\|_2$. One can use $\|\widehat{B}_\kappa - B^\star_\kappa\|_2 \le \|\widehat{\mathbf{B}} - \mathbf{B}^\star\|_F$, but this is loose by a factor $\sqrt{k}$. REAL-estimator allows to avoid this by showing that the error in $\widehat{\mathbf{B}} - \mathbf{B}^\star$ is spread roughly equally among different rows. *Regularity assumptions on context vectors:* we note that Assumptions 1-3 also appear in [6] and [15], and [6] discusseses and demonstrates their practical relevance on real data. Assumption 4 requires the data not to be heavy-tailed and explore all directions. Assumption 5 is concerned about the accuracy of low-rank estimator which is standard in low-rank matrix estimation literature.

[1]We adopt the same numbers for the references as the submission.