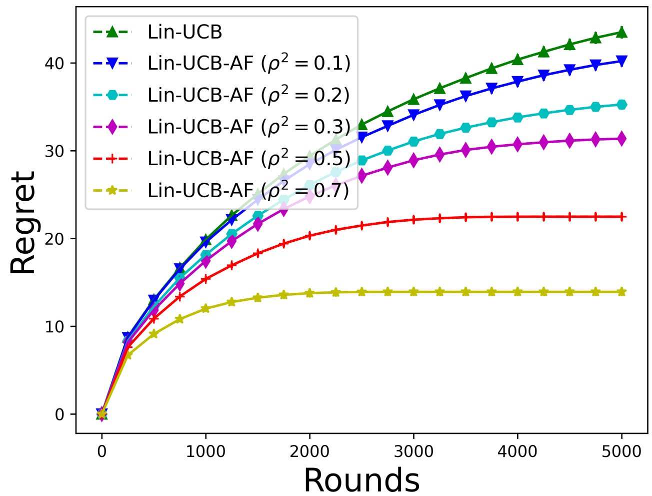

Figure 1: Regret of Lin-UCB-AF vs. varying correlation coefficients of reward and its auxiliary feedback. Here are the details of problem instance used for this experiment: as the variance of noise associated with is given by $\sigma^2 = \sigma_v^2 + \sigma_w^2$ and the correlation coefficient of $y_{t,a}$ and $w_{t,a}$ is $\rho = \sqrt{\sigma_w^2/(\sigma_v^2 + \sigma_w^2)} = \sqrt{\sigma_w^2/\sigma^2}$. To maintain the same noise variance across all instances, we set $\sigma^2 = 0.02$. We use $\rho = \{0.1, 0.2, 0.3, 0.5, 0.7\}$ and for each value of ρ, we first compute σ_w and then σ_v such that $\sigma_v^2 + \sigma_w^2 = 0.02$. As expected, the baseline bandit algorithm Lin-UCB performs worse and performance improved (smaller regret) as the correlation between reward and its auxiliary feedback increases.

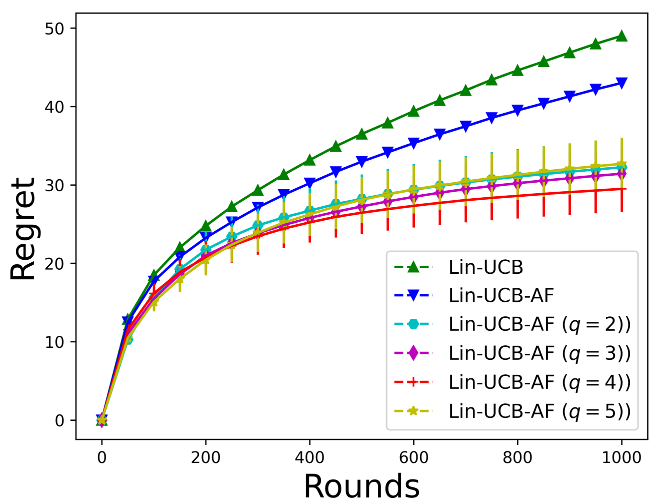

Figure 2: Regret of Lin-UCB vs. the varying number of auxiliary feedback when auxiliary functions are known. This experiment uses a bandit instance similar to the linear contextual bandit instance used in the paper but have a 6-dimensional synthetic contextual dataset and 5 auxiliary feedback functions with standard deviation $\{0.1, 0.8., 0.6, 0.4, 0.2\}$. As expected, there is a reduction in regret initially as q increases (auxiliary feedback with higher standard deviation is used first), but for $q = 5$, the performance declines (more regret compared to $1 < q < 5$). This result verifies that using more auxiliary feedback with estimated $\boldsymbol{\beta}$ may not always lead to variance reduction (Remark 1).