We thank the reviewers for their insightful comments and suggestions on our paper.

Reviewer 1: Thanks for pointing out these related papers. We discuss them below and plan that upon acceptance use the extra page to add these discussions to the paper as well.

• [Drusta WWW2017, ICML 2018]: These work focus on single buyer and a single product that is offered repeatedly. The (private) buyer's valuation of this product remains fixed across time. For case of strategic buyer, a discounted regret is considered with factor $\gamma < 1$ (similar to ours). An FES and PRRFES pricing policies are proposed which achieve the regret $O(\log\log T)$. However, their setting is very different than ours in that we consider 1) multi-buyer 2) feature based setting and the products offered at different rounds could be highly different. As a result, the buyers' valuations (that depend on the product features and also market noise) vary over time. For this we do not expect to have a policy with $O(\log\log T)$ regret in our setting. Closely related, Broder& Rusmevichientong (2012) consider the setting of a single truthful buyer and a retailer (single product) under a parametric choice model (corresponding to noisy valuation) and prove a lower bound of $O(\log T)$ for regret, using Van Trees inequality. One can follow that approach and via generalized Van Trees inequality prove a lower bound $O(d\log T)$ for the setting of CORP policy.

• Extension to nonlinear models: Although the paper focuses on linear valuation models, it is straightforward to generalize our analysis to some of the nonlinear valuation models. Specifically, consider model

$$v_{it}(x_t) = \psi(\langle \phi(x_t), \beta_i \rangle + z_{it}) \quad i \in [N], \quad t \geq 1\,,$$

where $\phi : \mathbb{R}^d \mapsto \mathbb{R}^d$ is a mapping and $\psi : \mathbb{R} \mapsto \mathbb{R}$ is an increasing function. Then by the change of variable $\tilde{v}_{it} = \psi^{-1}(v_{it})$, $\tilde{x}_t = \phi(x_t)$, we arrive at the relation $\tilde{v}_{it} = \langle \tilde{x}_t, \beta_i \rangle + z_{it}$. By modifying the CORP policy for this relation, we can get a policy that also achieve logarithmic regret for these nonlinear settings. Some examples include: log-log model, semi-log model and logistic model. While these models have been popular for some applications (the first two in hedonic pricing and the last in click-through-rate prediction), it is still an interesting direction to consider nonparametric models (e.g., Mao et. al. 2018 , Chen & Gallego 2018) but it is beyond the scope of current paper.

• With respect to Cohen et al. (and their tricks for robustness), their modified policy gets a regret of $O(d^2\log(\min\{T/d, 1/\delta\}) + d\delta T)$, where δ measures the noise magnitude: in case of bounded noise, δ represents the uniform bound on noise and in case of gaussian noise with variance σ^2, it is defined as $\delta = 2\sigma\sqrt{\log(T)}$. But then to have a logarithmic regret, δ should scale $O(\frac{\log T}{T})$, which we find very restrictive (not clear why the noise should shrink with time horizon and at a rate $1/T$). In comparison, we do not require such assumption.

• We will add "Conclusion" section in the revision and in our Related work section, we will add the following w.r.t contextual dynamic pricing with learning in non-strategic environment: Chen et al. (2015) studied this problem when the demand function follows the logit model and proposed an ML-based learning algorithm. Leme and Schneider (2018), Cohen et al. (2016), and Lobel et al. (2016) proposed a learning algorithm based on the binary search method when the demand function is linear and deterministic. In their models, buyers have homogenous preference vectors and are non-strategic. Hence, the problem reduces to a single buyer setting, where the buyer acts myopically, i.e., the buyer does not consider the impact of the current actions on the future prices. There is also a new line of literature that studied dynamic pricing with demand learning when the contextual information is high dimensional (but sparse); see Javanmard and Nazerzadeh (2019), Ban and Keskin (2017).

Reviewer 2: Thanks for your positive comments. We will fix the minor typos in the revision. For the lower bound please see the response to Reviewer 1.

Reviewer 3: (Discount factor:) To make the dependence of regret on the discount factor explicit, the regret bound (Theorem 4.1) works out at $O(\log(Td)\log(T) + \frac{\log^2(T)}{\log^2(1/\gamma))})$. Notably, the first term is due to the estimation error in preference vectors; that is, this term exists even if buyers were not strategic. The second term is due to the strategic behavior of the buyers and decreases as buyers get less patient (γ gets smaller).

Regarding the justification for discount factor: two types of targeting are common in online advertising (using HTTP cookies and using demographic information). While the former can potentially identify a unique user, the latter targets a group of users. We agree that for targeting based HTTP cookies, one can use time-varying contextual models. However, for targeting based on demographic information, using discount factors is more suitable. Specially that the number of users from these demographics is uncertain and the advertiser would like to keep showing the ads not to miss the opportunity of displaying his ad to a right user. The discount factor aims at capturing this opportunity cost. In cloud computing, a similar opportunity cost exists because many tasks are time sensitive.

• Footnote 5: We believe that one can extend our algorithms to a setting in which the seller has different objective functions such as weighted sum of buyers' welfare and his revenue or he aims at maximizing his revenue subject to return-on-investment (ROI) constraints. The difference is that the "optimal price" should be lower.