Reviewer 1:
Model assumption: We note that exponential sensitivity of the valuation with respect to covariate magnitudes can be resolved by using a logarithmic transformation of the covariates themselves. More generally, one may augment our approach with a machine learning algorithm which learns an appropriate transformation to fit the data well. Given these two observations the model is actually quite flexible, while admitting the learning guarantees we provide in the paper.

Assumption A3: Thanks for pointing that out. Indeed, that assumption is incorrectly restrictive – we apologize for the oversight on our part. Our results only need the following (weaker) assumption:

A 3 *Let $\theta^{(l)}$ be the l^{th} component of θ, i.e., $\theta = (\theta^{(l)} : 1 \le l \le d)$. We assume that there exist $\kappa_1, \kappa_2 > 0$ such that for each $z \in \mathcal{Z}$ and $\theta \in \Theta$ we have*

$$\kappa_1 \max\left\{(z^* - z)^2, \max_{1\le l\le d}(\theta_0^{(\ell)} - \theta^{(l)})^2\right\} \le r(z^*, \theta_0) - r(z, \theta) \le \frac{\kappa_2}{d+1}\|(z^* - z, \theta_0 - \theta)\|^2$$

where $\|(z, \theta)\|^2 = \left(z^2 + \sum_{l=1}^{d}(\theta^{(l)})^2\right)$.

Our results and intermediate lemmas remain unchanged with the replacement of assumption A3 in the paper with the above revised assumption. By making minor appropriate changes in the Appendix, such as in line 473 and Assumption A5 before line 538, the technical correctness of the paper remains intact.

Gamma: Indeed, in the definition of DEEP-C with Rounds in the Appendix, in line 392 we assume that $\gamma = \max\left(10\alpha_2^2, 4\frac{\kappa_2^2}{\log n}, \frac{\kappa_1^{-2}}{\log n}\right)$. We take your point and we will now mention this in the main paper. Note also that we discuss the role of γ as a hyper-parameter to be tuned in the simulation sections.

Side-information at higher prices: Indeed using the feedback of rejection at a price for higher prices causes biases as you suggest, which seems challenging to deal with. Also, it is not clear if the gain of using such a feedback is significant enough in our setting. Note that the optimal algorithm for stochastic model in Klenberg-Leighton (2003), which is special case of our model with $d = 0$, also does not use such feedback.

Reviewer 2:
Improving results for computational complexity and regret scaling w.r.t. d: Regarding computational complexity, our paper suggests an algorithm (Sparse DEEP-C) that is more computationally efficient than DEEP-C, and our empirical study shows promising results for this approach. Certainly further work on the value of sparsity remains an important future direction.

Lower bounds: We note that even though we assume a linear parametric model, the residual distribution is assumed to be non-parametric. While Kleinberg and Leighton (2003) solved the problem for the non-contextual setting in the presence of censured (binary) feedback, which is a natural and important assumption for the dynamic pricing problem, simultaneously learning the residual distribution and d dimensional parameters has remained an open problem, as evident from Table 1 in the paper. Our paper provides the first results at this level of generality. Also, as explained in our response to Reviewer 3, the structure of the problem does not meet the assumptions used in other general techniques such as bandits with side-information. While obtaining tight lower bounds on computational complexity and regret scaling w.r.t. number of dimensions d for our setup would certainly be an interesting challenge for future work, we believe our current results already represent a significant progress for the dynamic pricing problem.

Reviewer 3: Thank you for your positive assessment.

Computational complexity: See response to reviewer 2 above.

Relationship to bandits with side-information literature: Thanks for pointing this out. Indeed there are some similarities between our work and this literature, in particular in our work too the reward information is revealed for a subset of arms where the subset may be a function of the chosen action. We will note this point in the final version of our paper.

However, to our knowledge, each paper on bandits with side information assumes (a) a discrete set of arms, (b) the existence of a sequence of graphs indexed by time (possibly fixed) with the arms as its nodes, (c) the action involves pulling an arm, and (d) at each time the reward at each neighbor of the pulled arm is revealed. Our model does not satisfy this structure. For example, if we view a cell $(i, j_1, \ldots, j_d)$ as an arm, then the cell corresponds to a subset of prices/actions. The subset of arms for which the reward is revealed at time t depends on the covariate x_t, and the exact price p_t from the above subset. Thus, the assumption of a pre-defined graph structure is not satisfied. Alternatively, if one views each (z, θ) as an arm then the set of arms is uncountable, and it thus does not meet the assumption of a discrete set of nodes. Thus, as far as we can tell, the results from this literature are not directly transferable to our setting. Also, note that none of the prior works in dynamic pricing literature (e.g., see the references in Table 1 in the paper) assume that the set of user-valuations/prices is finite, since a constant error in price leads to linear regret.