We thank the reviewers for their time in reviewing our paper and their constructive feedback. We emphasize that the dynamic assortment selection problem we address in this paper is a fundamental and general problem – combining with contextual information and an easy extension to position-based offering (see the response to Reviewer 3) makes our setting one of the most common forms of recommendations one may face in practice. We propose the methods which are both theoretically sound and practical. The responses to specific questions/issues raised by each reviewer are presented below.

Response to Reviewer 1

The parameter θ^* is modeled from the entire data considered and then in each round the MNL model is simulated with this θ^*. We will make it more clear in the paper.

We are currently working on the experiments with model mis-specification where the true model is not the MNL model. We will include the results in the paper. We appreciate your input.

Response to Reviewer 2

We agree with the suggested corrections and appreciate your feedback. On Assumption 1, while it states $x_{t,i}$ is i.i.d., we emphasize that the i.i.d. assumption on $x_{t,i}$ is only required during the initialization phase to ensure the invertibility of V_{T_0} (in order to have a unique solution of MLE as mentioned in Appendix B.1). After the initialization, $x_{t,i}$ can even be chosen adversarially as long as $\|x_{t,i}\|$ is bounded. Now, we note that in practice the same can be achieved by introducing regularization, but for a better tractability of the analysis we chose random initialization along with the i.i.d. assumption (at least for the initialization). We will make it more clear in the paper.

As for the experiments, we will include more details about the experimental setup in the main body of the paper if the space permits. On the optimization step of the experiment, MovieLens dataset does not contain different revenue value for movies. Therefore, it is equivalent to having the unit revenue for all movies. Hence, the optimization step reduces to a sorting task based on estimated utility. We will include additional experimental results where we use synthetic data which contains the (synthetic) revenue parameter for each item. For that, we use the LP solution proposed in [17].

The experiments are indeed semi-synthetic. The interactive aspect of bandit problems (not just our MNL bandit) makes it notoriously difficult to evaluate in real-world settings unless one performs a field experiment. That is why most bandit papers perform evaluations with completely synthetic data. As mentioned in the previous paragraph, we will include additional experiment results with synthetically generated data.

Response to Reviewer 3

Our work (as well as previous work in MNL bandit) is distinct from other combinatorial bandit problems such as cascading bandits and semi-bandits. In typical cascading or semi-bandit settings, the mapping from the item context to the user feedback is independent of other items in an offered set. On the other hand, MNL choice feedback is a function of entire assortment which makes our analysis much more challenging.

Regarding position-based offering, we can easily incorporate display position effect within the assortment by including a categorical variable indicating the display position in the context vector. Hence, we also estimate parameters corresponding to each display position. We can show that our algorithms are able to still use the LP solution [17] for this position-dependent extension of the combinatorial optimization problem. Note that this extension is different from previous position-based click models in which the user feedback is typically a function of an item context independent of other items in an offered set. This position-based extension of our framework still takes into account the substitution effect within the assortment.

We argue that in practice the assortment based offering (with or without display position effect) is the most prevalent form of recommendations in online retailing (e.g. Amazon, Walmart), streaming services (e.g. Netflix), news websites, and many more – in fact, one rarely faces single-item offering (typical bandit setting) or item-wise cascading offering in those common applications. Furthermore, when an assortment (a set of items) is offered, there is often a substitution effect among the items, which many other combinatorial bandit models do not address. We appreciate your feedback and the chance to re-emphasize our motivation.

We will add additional experimental results on the computation time. Regarding the evaluation on real-world data, (as mentioned in the response to Reviewer 2) the interactive aspect of bandit problems (not just our MNL bandit) makes it notoriously difficult to evaluate in real-world settings unless one performs a field experiment.

Thank you for pointing out the citation mistake. We replaced the arXiv version of [39] with its UAI 2016 publication. We double-checked the other arXiv papers which we cited in our paper and confirmed that they did not appear in any previous proceedings or journals.