We want to express our gratitude to all the reviewers for careful reading and valuable comments. We believe the advice will help us bring a better version of this paper. To begin with, we want to apologize for the typos and unclear writings. We will correct them in the final version, and add the broader impact section. It appears that there is a major concern regarding the contributions of our paper comparing with two lines of previous work, which we make clear first.

The first line of research is the various (unbiased) propensity estimation methods in the recommendation literature. The papers mentioned by the reviewers all assumes on a click model: $p(\text{click} = 1|x) = p(\text{expose} = 1|x) \cdot p(\text{relevance} = 1|x)$. Note that the implicit assumption being made is: $p(\text{expose} = 1, \text{relevance} = 1|x) = p(\text{expose} = 1|x) \cdot p(\text{relevance} = 1|x, \text{expose} = 1) = p(\text{expose} = 1|x) \cdot p(\text{relevance} = 1|x)$, so $p(\text{relevance} = 1|x, \text{expose} = 1) = p(\text{relevance} = 1|x)$, suggesting that relevance $\perp$ exposure $|\, x$, i.e. relevance is independent of exposure given the features. This may not be true (or at least cannot be examined) in many scenarios, unless we are able to collect every single factor that may affect the users' decision making process into x. The merit of our approach is that we get rid of the dependency on the click model assumption, and provide an alternative solution for researchers and practitioners who suspect the validity of relevance $\perp$ exposure $|\, x$ in their data. In practice, the gain from our approach depends on the degree of violation on the above assumption. Therefore, compare with the prior work, we introduce new perspectives and a feasible solution to this challenging problem. As for empirical evaluations, we have also tried our best to add more baselines according to the reviewers' requests (Table R.0). We ran into some trouble replicating the baselines with the published implementation, but we keep the our implementation as consistent as possible with the original work.

The other line of prior work is distribution-robust optimization (DRO), which is a vast domain. While our model also belongs to this category, the critical component that we argue for robustness is the propensity score distribution, which to the best our knowledge has not been studied before. The majority of papers in this domain have a different emphasis on the robustness of feature distribution or data generating distribution, which do not apply to our problem since the propensity score does not have a generative nature. From a technical perspective, the challenging part is that the propensity score term is placed on the denominator, so it requires extra proofs and arguments to obtain the duality, relaxation and concentration results. Our contribution is also novel in this regard.

To Reviewer#1. We thank Reviewer#1 for the insightful questions. We compare our work with other DRO and unbiased propensity estimation method as above and provide empirical comparisons with the mentioned baseline, where our approach still shows better performance. The reason why POP only give minor improvements in simulation is a consequence of the simulation setup where popularity is not directly related to exposure. The ORACLE methods may experience fluctuation because we have added random noise on the oracle to simulate the data, so ORACLE is only an unbiased estimation, but the variance can be large. Finally, we agree that the benefit of our approach is less significant when having access to the exposure strategy (which would be ideal), but this rarely happens in reality.

To Reviewer#2. We thank Reviewer#2 for pointing out the insufficiency of our manuscript, and we provide a refined analysis on the prior literature, including their weakness and comparisons with our work, in the above paragraph two. As we mentioned, uncertainty in exposure has not been well-handled by the unbiased propensity estimation methods, since they rely on another implicit assumption (the assumption for the click model) that may not be correct. Our approach provides an alternative solution that is free from the assumption. As for the empirical evaluations, we managed to add one set of additional experiment for the suggested baselines. It is possible that we have not tuned the baselines to perfection, but based on the initial result, the proposed approach still outperforms the propensity estimation approaches.

To Reviewer#3. We thank Reviewer#3 for the suggestions on further improvements. We apologize for the inconsistent scale in Table 1 where we forgot to multiply by 100 on Goodreads' results. Here, we provide a more detailed comparison with the mentioned literature in the beginning and add one set of experiment to include the SOTA baselines, where our approach still outperforms the propensity estimation approaches. As we mentioned in our response to Reviewer#1, the inconsistent performance of POP and ORACLE in simulation is a consequence of how we generate the data.

To Reviewer#4. We thank Reviewer#4 for the advice on providing a more in-depth comparison with the counterfactual recommendation literature and include more SOTA methods as baselines, which we provide at the beginning of this rebuttal. We wish to point out that the primary focus of recommendation is on the supervised learning part. Although we introduce a counterfactual learning component, the final model should still be examined in the supervised learning setting. The analysis for the identifiability issue, on the other hand, is a heated topic for the sensitivity analysis and would require another research paper to explore under the recommendation setting, which we pursue as future work.

	MovieLens-1M simulation data					**MovieLens-1M real data**				
	URL-MF*	LtR*	ExpoMF	DM	ACL-MLP	URL-MF*	LtR*	ExpoMF	DM	ACL-GMF
Hit@10	16.33(.2)	19.31(.4)	16.26(.9)	16.99(.8)	**21.58**(.1)	63.71(.2)	64.24(.1)	62.50(.7)	63.33(.8)	**64.32**(.2)
NDCG@10	7.26(.3)	7.91(.3)	7.24(.6)	7.47(.5)	**8.42**(.2)	33.19(.2)	33.43(.1)	32.85(.4)	32.97(.5)	**33.70**(.1)

Table R.0: Extra results on Movielens-1m simulation and real-world data. **URL-MF***: *Unbiased Recommender Learning from Missing-not-at-Random Implicit Feedback, WSDM'20* (the published code is not executable); **LtR***: *Unbiased Learning to Rank with Unbiased Propensity Estimation, SIGIR'18* (the published code is for search ranking); **ExpoMF**: *Modelling User Exposure in Recommendation, WWW'16*; **DM**: *Modelling Dynamic Missingness of Implicit feedback for Reommendation, NeurIPS'18*; **ACL-X**: the proposed adversarial counterfactual approach with model X as f_θ and g_ψ. Results have been multiplied by 100.