We appreciate the reviewer's constructive suggestions. Reviewer 1 and 2 highly evaluate our work. It looks the score of Reviewer 3 and 4 is mostly due to our paper's presentation style. We will sincerely take their suggestion into account in a revision phase. For each question, we answer as follows.

Reviewer 1 **Weakness:** We will try to add more insight into the theoretical approach, such as the difference between our case and non-covariate shift case. **Additional feedback:** (1) Using an oracle behavior policy is not enough to achieve the efficiency bound because we need a control variate. Estimating behavior policy in an unknown case is seen as using some control variate. We will add this explanation and improve the experiment; (2) Cross fitting versions of DM and IPW estimators do not have an efficiency property (Theorem 2) since these estimators do not have the doubly robust structure. We will also add the explanation of Donsker conditions more following your suggestion.

Reviewer 2 **Additional feedback**: (1) This is a good point. Unlike Theorem 1, we cannot obtain the explicit form of the efficiency bound under this i.i.d sampling; (2) Exactly. We will add this point; (3) Yes, we would follow your suggestions.

Reviewer 3 **Weakness**: (1) Due to the space constraint, we could not add a full explanation for each theoretical result. We will introduce a more intuitive explanation by rearranging the structure (we can make more space following Reviewer 1 and 2 suggestions); (2) Although an idea of IS itself is widely known, the analysis of IS based estimators is specific to each setting. We emphasize that this analysis is much more than the extension of this literature. Note that we cited covariate shift literature in the Introduction and Preliminaries. Here are the differences of Shimodaira (2000); Sugiyama et al. (2008) and ours. First, the goal is different. Our goals are OPE and OPL. On the other hand, their goal is solving regression problems or evaluating the expected log-likelihood. Second, due to the difference in goals, the analysis of the estimator is completely different. In OPE, we calculate the asymptotic first-order term of the estimator when plugging nonparametric estimators into the density ratio (Theorem 2). This implies that our analysis takes the plug-in effect into account. The general density ratio estimation literature such as Sugiyama et al. (2008); Shimodaira (2000) does not analyze this type of plug-in estimator though they actually use in practice. The effect of the plug-in is not negligible in OPE since the asymptotic variance is generally changed due to the plug-in. In this sense, our analysis is considered to be more sophisticated and tailored to an OPE problem. Third, for our OPE setting, we show not only the asymptotic distribution of estimators but also the efficiency bound. (3) For OPL, we agree with your opinion. We will incorporate it into the experimental section. On the other hand, in OPE, to estimate the policy value, we need to conduct covariate shift adaptation as our paper. **Relation to prior work**: This is due to space constraint since we try to refer to various literature as much as possible rather than explaining specific literature in detail. We will try to add more explanation. **Additional feedback**: We will add references in the parts you mentioned. The meaning of "Note that we can ..." is the i.i.d case where one observes $S \sim \mathrm{Bern}(\rho)$, along with X when $S = 1$ and Z when $S = 0$.

Reviewer 4 **Weakness**: (1) We appreciate the reviewer's detailed suggestion. We would try to make our contribution more clear following it; (2) We agree, but we also think this problem is a more general OPE problem rather than ours; (3) We will cite Sondhi, et al. (Note we already cited Barenboim & Pearl's work). (4) In Remark 2, we discussed a related technical difference between the shift of action and covariate. For the method, we do not need to estimate the density ratio simultaneously. For example, in the case where there are two actions $\{1, 2\}$, we estimate $p(a = 1 \mid x) = \frac{p(a=2)p(x|a=2)}{p(x)}$ and $\frac{q(x)}{p(x)}$. We can obtain $p(a = 2 \mid x)$ by $1 - p(a = 1 \mid x)$. If estimating the density ratio directly, we estimate $\frac{p(a=1)p(a=1|x)}{q(x)}$ and $\frac{p(a=2)p(x|a=2)}{q(x)}$. Thus, in both cases, **we need two estimators**. In addition, in general, **there is no significant difference between the convergence rates of** $p(a = 1 \mid x)\frac{q(x)}{p(x)}$ **and** $\frac{p(a=1)p(x|a=1)}{q(x)}$**.**. Moreover, before submission, we tried a method based on simultaneous estimation of the density ratio, but the performance was not good. We consider that the density ratio estimation is harder to fit nonparametrically than logistic regression. As a result, for ease of presentation, we did not adopt such a direction; (5) The purpose of rejection sampling step of Dudík et al. (2014) is to stabilize an estimator constructed from a dynamic and nonstationary policy, where $\pi^{\mathrm{e}}/\pi^{\mathrm{b}}$ can take an extreme value. In such a case, the technique will stabilize the estimator heuristically. However, there is a trade-off between bias and stabilization. Hence, in general, we do not have to use heuristics such as rejection sampling and clipping for all cases. (6) Firstly, we showed the experimental result using a direct density ratio estimator (IPWCS-R) as mentioned in L290. As you mentioned, it does not work in practice. Secondly, We introduced IPW-type estimators as a meta-estimator allowing any types of methods to estimate the density ratio. We analyze the specific IPW estimator when using the Nadarya-Watson estimator to estimate the density ratio since the asymptotic analysis is relatively straightforward. (7) We will investigate it. (8) As mentioned in L268, for OPE, we showed the experimental results with different sample sizes, 300, 500, 1000, in Appendix. We will add the results of OPE and OPL with more sample sizes in the next revision. **Clarity**: As far as our knowledge, y is also used. I suppose the difference of this convention is due to the difference of communities, causal community, and RL community.