

1 We are grateful to the reviewers for the valuable comments. Below we respond to each and every point of the reviewers.

2 **Reviewer 1:** We thank the reviewer for the comments. Please find our point-to-point response below.

3 *Propensity score:* Our work does consider a mild constraint on the propensity score (encoded by the likelihood ratio
4 κ in Assumption 2), and our minimax risks also have tight dependence on κ . We agree that other modelings of the
5 propensity score are also possible (e.g. modeled as another nonparametric function), but for size constraint we decide to
6 leave it as future work.

7 *Guideline on algorithm design:* While we agree that the main focus of our work is to characterize the tight minimax rates
8 in theory, we also provide several practical insights. First, estimating two baselines separately and differencing is no
9 longer optimal when the HTE is smoother, and one should estimate the HTE directly based on the pseudo-observations
10 obtained via covariate matching. Second, the covariate geometry plays a central role in HTE estimation, and one should
11 discard observations with poor matching quality. Their usefulness is illustrated in the empirical results below.

12 *Comparison with Alaa and Schaar'18:* As correctly claimed by the reviewer, the main difference is on the assumption:
13 they imposed smoothness constraints on two baselines, while we assume a smoother HTE. This difference leads to
14 several consequences. First, one of their main take-aways is that the minimax rate is determined by the less smooth
15 baseline, while in our case it is determined by an interpolation of both smoothness parameters depending on the
16 covariate geometry. Second, we also provide new practical guidelines to employ the covariate geometry and perform
17 covariate matching (see above), a point which is unnecessary in [AS18] where the baselines are estimated separately.
18 We cited this work in Line 121-123 of the submission, and will elaborate more on the differences in the final paper.

19 *Line 295-296:* Thanks for pointing out this statement which is not very precise. In our work, we study the minimax risk
20 after conditioning on both random realizations of the covariates and group assignments, meaning that the worst-case
21 HTE can depend on both the above realizations. In contrast, [NW17] only conditions on the random covariates and
22 considers an expected risk taken with respect to the randomness in the group assignments, where the worst-case HTE
23 cannot depend on the group assignments. We will further clarify this point in the final paper.

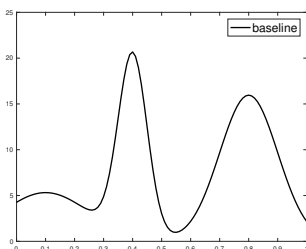
24 *Assumption 1:* This assumption is a nonparametric modeling of the fact/belief that the HTE is simpler than the baselines.
25 For example, in (semi)parametric modelings the HTE is typically assumed to be a constant or a low-dimensional/sparse
26 linear function; our assumption is its nonparametric counterpart. We agree that we could also assume the HTE to be
27 approximated by a smoother function, but without additional assumptions, this approximation error will enter the play
28 as an additive error after a simple use of the triangle inequality. So for brevity we assume the HTE itself is smoother.

29 **Reviewer 2:** We thank the reviewer for the comments. Please find our point-to-point response below.

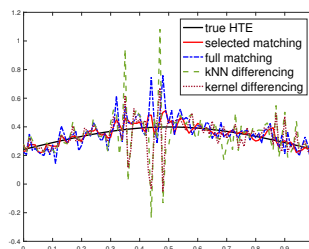
30 *Fixed grid and smoothness constraint:* We agree that the fixed design assumption is not very practical, and we include it
31 mainly to illustrate how the combination of the smoothness and the covariate geometry determines the final minimax
32 rate (cf. Line 62-63 of the submission). We also agree that restricting to smoothness less than 1 is unfortunate for us in
33 theory, but we could provide three reasons for this. First, capturing higher order smoothness could be a challenging task
34 in nonparametric problems, and there are a number of other examples where how to solve this task is open. Second,
35 there are insights suggesting that the cases $\beta > 1$ and $\beta < 1$ exhibit fundamentally different behaviors in the minimax
36 rate. Third, there seems to be a fundamental approximation-theoretic difficulty when arguing the lower bounds for
37 $\beta > 1$. We refer to Line 267-287 of the submission for details.

38 *Well-known machinery:* We agree that the kernel or nearest-neighbor tools are well-known in nonparametric methods.
39 Nevertheless, our main aim is to provide a careful combination of these tools which also takes the covariate geometry
40 into account, and show that the aforementioned combination (nearly) attains the minimax rates for HTE estimation.

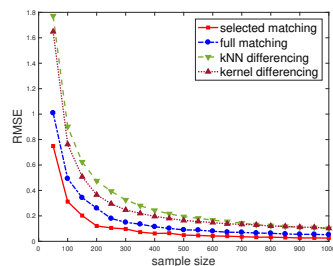
41 *Empirical results:* Thank you (and also Reviewers 3-4) for bringing up this useful suggestion. We will add an extensive
42 collection of numerical experiments to the final paper, and we plot one example here to illustrate the performance
43 of our algorithm. Here (a) plots an oscillating baseline function μ_0 with $d = 1$, and (b) plots a smoother HTE τ .
44 We compare 4 estimators: *selected matching* as used in the paper, *full matching* without discarding any observation,
45 *kNN differencing* and *kernel differencing* which apply different methods to estimate both baselines and then take the
46 difference. Figure (b) plots one random instance of all estimators with sample size $n = 1,000$, and (c) plots the root
47 mean squared error (RMSE) as a function of sample sizes. We observe that utilizing covariate matching performs better
48 than estimate-then-difference, and throwing away covariates with poor matching quality also leads to a smaller error.



(a) Baseline function μ_0 .



(b) HTE τ .



(c) RMSE vs sample size.

49 **Reviewer 3:** We thank the reviewer for the comments. We will add numerical experiments and results to the final paper
50 to illustrate the empirical performance and practical implications; please refer to our response to Reviewer 2 above.

51 **Reviewer 4:** We thank the reviewer for the comments. We will add numerical experiments and results to the final paper
52 to illustrate the algorithms in action; please refer to our response to Reviewer 2 above.

53 *Response to minor comments:* The complete statements of Theorems 3-6 are in the main document, and only the proofs
54 are relegated to the appendix. Meanwhile, we also provide the high-level proof ideas for these statements in the main
55 document. We will also replace the dot notation by tilde notation to gain more visibility - thanks for pointing out this.