

1 We thank the reviewers for taking the time to read our manuscript carefully, as well as for providing very insightful
2 and constructive comments. We are flattered by their mostly positive feedback (e.g., “neat idea!”, “excellent empirical
3 performance”, “elegant solution”, “potential to be useful”, “mostly well written”) and we welcome their suggestions on
4 how to improve the writing to make the paper more accessible. We respond to the reviewers’ comments in detail below.
5 We feel that the suggested improvements are relatively few and will be easy to implement in a minor revision.

6 **Extensions**

7 *Label-conditional coverage.* To answer the reviewer’s interesting comment: yes, provable label-conditional coverage
8 can be easily achieved by calibrating the threshold τ separately for each class. More precisely, focusing on the exten-
9 sion of Algorithm 1 for simplicity, we would compute $\hat{Q}_{1-\alpha}^{(y)}(\{E_i\}_{i \in \mathcal{I}_2})$ as the $\lceil (1-\alpha)(1 + |\{i \in \mathcal{I}_2 : Y_i = y\}|) \rceil$ th
10 largest value in $\{E_i\}_{i \in \mathcal{I}_2: Y_i = y}$, for each $y \in \mathcal{Y}$. Then, we would define $\hat{\tau} = \max_{y \in \mathcal{Y}} \hat{Q}_{1-\alpha}^{(y)}(\{E_i\}_{i \in \mathcal{I}_2})$ and output
11 $\hat{C}_{n,\alpha}^{\text{sc}}(X_{n+1}) = \mathcal{S}(X_{n+1}, U_{n+1}; \hat{\pi}, \hat{\tau})$. We would be happy to include this extension in a revised manuscript.

12 **Minor comments**

13 *Size of data splits for CQC.* We agree with the reviewer that this point should be clarified. For simplicity, we split the
14 data into subsets of equal size for all methods, including ours (Section 2.2). No effort was made to optimize the size of
15 the splits for any method, so the empirical comparisons are fair. We do not expect the results of our experiments to
16 change meaningfully if the sizes of the sample splits are optimized, since our method has the advantage of requiring
17 one fewer split, and it has a stronger optimality property in theory.

18 *Relation with prior work.* As highlighted by the reviewer, it is clear that the prediction sets in (5) and the conformity
19 scores in (7) are the novel contributions of our paper. Our method indeed builds upon the theory on model-free predictive
20 inference previously developed by others, and we gratefully acknowledge those works. That said, to make conformal
21 inference useful, one needs high-quality conformity scores leading to tight prediction sets, which is the focus of this
22 paper and the current research frontier.

23 *Choice of data sets.* The reviewer correctly points out that our data sets are “standard” and in that sense “quite boring”.
24 However, we feel that our choice is well-justified because our goal is to make the results easily accessible to the largest
25 possible audience without distracting from the methodological message of the paper. Furthermore, choosing an unusual
26 data set for comparing methods, without a good reason, may give the wrong impression that the example is somehow
27 “cherry-picked”.

28 *Independence vs. exchangeability.* Our wording when we said “independence is unnecessary” was accidentally a little
29 ambiguous. What we meant is that “exchangeability” is sufficient. Our assumptions are stated explicitly later in the
30 paper, but we acknowledge that the sentence “independence is unnecessary” at the beginning should be clarified. We
31 thank the reviewer for bringing this issue to our attention.

32 *Labels with zero probability.* The referee is right: our comment about zero-probability elements can be removed. We
33 thank the reviewer for suggesting this improvement.

34 **Exposition**

35 *Definition of the generalized inverse \mathcal{S} .* We would be happy to provide some intuition for this definition before the
36 formal statement, following the referee’s suggestion.

37 *Definition of E in (7).* We would gladly follow the referee’s suggestion to describe E in words with something along
38 the lines of “the smallest probability τ s.t. y is a member of the generalized inverse \mathcal{S} ”.

39 *Referencing (8) in the text.* We agree that (8) should be referenced more clearly in the text; we will gladly do so in the
40 revised manuscript.