We thank the reviewers for their helpful comments. We first provide individual responses to each reviewer's comments and then provide a general response to all the reviewers.

Reviewer 1: Tuning of hyperparameters must be done with cross validation or there is no way to ensure that the tuning is not causing over-fitting. If we tune for maximum accuracy and test on the target data separately the models perform very poorly. Handling this overfitting is why the uLSIF and KLIEP papers were written, the algorithms of which rely on finding the right parameters for optimal reweighting using cross-validation.

Reviewer 3: The paper presents a novel approach to reweighting data in the covariate shift which unlike other state of the art methods doesn't require parameter tuning and shows promising signs of being more stable. Further work can, and will, expand on how these ideas can be scaled to larger datasets and larger dimensions.

We focus on the SVM because it enables us to provide the most straightforward demonstration of our framework on the real datasets used by other state of the art covariate shift method papers. However, the method is model-independent and can be applied on other datasets where the underlying target function can be better modeled by other model classes like NN or decision tree based GdBt. We also want to point out that SVM might be old but are used heavily in practice when the data is noisy and limited (for example in finance applications). Further, there are many options that can be explored as future work. For example, one could easily apply this method to the last row of a neural network. The NN papers mentioned by the reviewer address domain adaptation while this work addresses the covariate shift problem (though they are qualitatively similar, they differ mathematically).

Reviewer 4: We will make the suggested changes to improve the writing. We will correct the typo in Equation 1 and replace $L(f(X), X)$ with $L(f(X), y)$ and define the notation clearly. Fig 1c in L57 is similar to Fig. 4a explained in experiment 3.3 except it uses 100 training points instead of 200. We will make this reference in the text of the paper. The reviewer correctly points out that [19] doesn't estimate individual densities but directly estimates the weight. We will make the correction. We have tried to distinguish between techniques that estimate individual densities via KDE against directly estimating the ratio of densities in the subsequent paragraph where we mention the methods by Huang et al., KLIEP and uLSIF. We will further clarify those differences. We appreciate you bringing this distinction of terminology. However, it is also important to realize that all of these previous methods still require a *kernel function* with a properly *tuned bandwidth* - which is one of the main shortcoming that our paper addresses.

The shortcomings of our method amount to it needing further work to handle higher dimensions . We explain one way around the instabilities of the CDF in higher dimensions by a heuristic (the "additive" V-matrix), but we believe a rigorous treatment of this will be the subject of additional papers (this could possibly be a reason for worse performance on twonorm dataset).

The paper addresses the hyper-parameter tuning problem that state of the art covariate shift methods face for *finding optimal reweighting.* In the context of our paper, all reweighting is encoded in the V-matrix, a concept introduced by Vapnik et. al that we demonstrate includes all previous methods as a special case. The regularization parameter γ in Equation 4 is a learning algorithm parameter, which is not involved in calculating the V-matrix reweighting. Previous methods can be interpreted as calculating diagonal V-matrices for reweighting, and require tuning in order to do this. Similarly, the kernel width for f in V-SVM is a model (learning algorithm) parameter and again should not be confused with parameters that are needed to be necessarily tuned for optimal reweighting. $\gamma = 0.1$ and the kernel width for f is chosen to be 1 for all the methods tested in Experiment 3.3.

Reviewer 5: We agree the title is non-informative, but we felt it was necessary to emphasize the main contribution of our work since we wanted to emphasize on two key aspects of our method - robustness and no parameter tuning. Perhaps a better title would be: "Pairwise covariate-shift reweighting from cumulative distribution functions gives robust parameter-free performance"?

All Reviewers: In this paper we rethink how we find the optimal reweighting for the covariate shift problems with an emphasis on removing the necessary parameter tuning and increasing the robustness. As we emphasize in our paper, this framework is model-independent. We define a loss function - so any hypothesis (model) class that minimizes a loss function will work. To this end along with giving certain theoretical guarantees for our method, we demonstrate by experiments on synthetic and real datasets that our method gives near similar or improved performance than other methods in most of the cases *without any need for parameter tuning* that the other covariate shift methods heavily rely on for finding the optimal reweighting. *In addition to that, for other covariate shift methods, there is no rigorous way to tune the right parameters. Hence, a method that can avoid the need for parameter tuning without any compromise on the performance (which other methods only achieve if the right set of parameters are tuned) is a novel contribution.* Further, our method shows *stability* in its predicted probability function (shown in our synthetic data experiments), and *consistency* across multiple real datasets. We would like to point out that the state of the art methods we tested, do not provide large improvements from the unweighted case in some experiments. These state of the art methods sometimes do significantly worse than the unweighted cases if the parameters are sub-optimally tuned. The bold face in Table 1 is to emphasize the consistency of our method demonstrating that our method gives the best mean performance for *6 out of 12 experiments* shown in Table 1. While we certainly can add more experiments, we believe the current number (12 real data experiments and 1 synthetic data experiment) is enough to justify our claim. The nature of experiments performed in the paper is an established standard for evaluation in the popular covariate shift papers (uLSIF, Huang et al.).