We thank all reviewers for thorough and constructive feedback. Due to space constraints, we only discuss major concerns here and will incorporate all of your concerns in our camera-ready version.
When we refer to **Reviewer #n, m within the text, we mean *our response to that point*, not the original review.**

Reviewer #1, [3.1] We appreciate and are grateful for the overall positive stance on our submission. We agree that the main contribution is to provide a shift of formulation of an existing concept with solid theoretical foundations, but we believe novel ideas and applications can often stem from such new formulations that could not have been born from the existing approaches (e.g. see **Reviewer #2, 4**). We thus believe that our contribution is well-aligned with the NeurIPS evaluation criteria, which seek "[s]olid, technical papers that explore new territory or point out new directions of research [...]". We also believe there are potential contributions made on the empirical side – please see **[6.3]** below.
[3.3] The key ideas in the proof are Chebyshev's inequality and the stability notion in [Bousquet and Elysseeff, 2002], which are by now standard techniques in learning theory. The main difference is that operator-based approaches are not based on vv-regression, but rather the theory of linear operators, which does not facilitate the use of such techniques.
[3.4] We propose to run some experiments to verify Theorem 4.4 and include the results in the camera-ready version.
[3.5] This is a valid point. The exact expression for the convergence rate are expressed on page 22/23 of the Appendix as (†), (††) and († † †). We propose to include this as part of Theorem 4.4 in the camera-ready version.

[6.3] The contributions over [Grünewälder et al., 2012] (hereafter written [G]) are as follows. We apologise for not making them clear in the submission and we propose to make more explicit in the camera-ready version. (i) We consider the CME as an explicit function $\mathcal{Z} \to \mathcal{H}_\mathcal{X}$, which we believe is a more principled way of motivating regression, as opposed to [G] who apply the Riesz representation theorem for each $Z = z$ and obtains an objective function which is not in the form usually seen in regression [G, eq. (5)]. Although this is equivalent to our squared-loss, and therefore leads to the same empirical estimates, our approach facilitates the use of other loss functions more easily, as well as more complex kernels, which will lead to different empirical estimates; please see **Reviewer #2, 2, 3**. (ii) We believe our theoretical analysis is more thorough than [G], as we derive a new result for convergence rate and provide new analysis of what the surrogate loss means exactly (L266-L286), which we believe is more complete than [G, Thm 3.2].

[5.3] This is a valid point; we agree that the paper would read better if we move (at least parts of) Appendix B into the main body of the paper. This was purely due to space constraints. We propose to do so in the camera-ready version.
[5.4] This is a good point – we propose to mention the uncentred case in the camera-ready version, since they are often used in the literature. The advantages of our CME approach remain true against the uncentred case.

Reviewer #2 We thank the reviewer for the overall positive view. For the first concern, please see **Reviewer #1, [3.4]**.
2, 3. This is a great comment. It is true that, despite its simplicity and intuitiveness, the squared-loss does come with some disadvantages, particularly in terms of robustness. By using a different loss on L219, our new formulation facilitates the use of other loss functions conveniently for a more robust RKHS representation of conditional distributions, and we believe [Laforgue et al., 2019] would be a relevant paper to cite. Moreover, on L221, it is not necessary to use a scalar kernel composed with identity; any other operator-valued kernel can be used, and different empirical estimates will be obtained than the closed-form in (3). This ability to facilitate other loss functions and OVKs is a definite advantage of our new formulation over the existing approach, and we propose to emphasise this point in the camera-ready version. The subsequent closed-form empirical estimates and theoretical analysis that follow are based on the squared-loss and scalar kernel composed with identity, and we propose to keep these as they are, since this case is the most basic and common case and we believe the results are therefore still of value.
4. This is a great point, and we have actually been working precisely in this direction, including developing this into a statistical test. Expectation over Z would be a good way of aggregating the random criterion into a single number, and indeed we can also apply it to HSCIC, but there are also other ways of doing it, e.g., considering the vv-RKHS norm of the difference of the estimations. We argue that this is an example of how our new formulation of CMEs can open up new questions and applications. In this paper, we left it random to retain the direct analogy with the unconditional case.
5. As long as the conditioning variables are absolutely continuous with respect to each other, they are not required to be the same. We propose to let the conditioning variables be different in the camera-ready version, for full generality.

Reviewer #3 We thank the reviewer for a critical review. For the main concern over practicality, we believe the shift of formulation we propose is such that it enables new applications that were not previously possible. A concrete example is laid out in **Reviewer #2, 4**, and a real-world application could be the estimation of the distributional treatment effect, comparing P(Y|X,T=0) and P(Y|X,T=1) where Y, X, and T denote outcome, covariate, and treatment variables, respectively. This is an important problem in medicine, public policy, and economics. The CME was first introduced in the machine learning community (Song et al., ICML2009) and most important papers in this area have been published in ICML, NeurIPS, AISTATS, etc. Hence, we believe NeurIPS is the right venue (see also **Reviewer #1, [3.1]**). Also, we will further improve clarity in our camera-ready version (in particular, please see **Reviewer #1, [5.3]**).

Reviewer #4 We thank the reviewer for the positive view of our paper. As for your concern, unfortunately, the previous stronger assumptions are virtually always violated. For the relevant discussion and references, please see L131-147.