We thank all reviewers for their time and comments. Below, we try to address some of the main questions raised by each reviewer, and will edit the paper to include these additions.

Reviewer 1

1. **Comparison of ACNet with other neural density estimators.** We consider ACNet to have significantly different goals from flows and other density estimation techniques. ACNet learns Copulas rather than joint densities. It allows us to compute and backpropagate through quantities such as probabilities, conditional cumulants etc., on top of densities (Sec. 3.2 and Appx. 7.6). This has numerous applications such as the uncertain data setting we experiment on (Sec. 4.3). In contrast, common neural density estimators require numerical integration to compute these quantities and may face significant numerical challenges especially when backpropagating through them. *ACNet should be considered as a neural method for learning Copula/joint CDFs, rather than an alternative for flows and other neural density estimators*.

2. **Computational cost.** Inversion typically requires no more than 50 forward-backward passes of ACNet (Sec. 3.1). This can be regarded as a 'constant factor', though its effect is clearly non-negligible. Training is naturally slower than the parametric copula we compare to, since ACNet possesses many more parameters. We do want to point out the following. (a) Our method is able to learn Copula which parametric copula were unable to learn (closing prices of GOOG-FB), and as such, should not be simply treated as a basket of existing parametric copula. (b) In all experiments, ACNet took no more than 2 hours with a standard consumer laptop (Sec. 4). (c) The number of inverse calculations required grows linearly with the number of dimensions—this follows from the fact that the inversions are of 1-dimensional functions. (d) Inversions are done by performing Newton's method in a 'vectorized' manner. This results in very significant speedups in practice.

3. **Evaluation of ACNet using existing benchmarks for density estimation.** We choose finance as a target domain since this has been the traditional application of Copulas. Nonetheless, we agree that it would be good to apply them to the simple datasets such as GAS and POWER and will include it in the final version of the paper.

4. **Experimental setup.** The normalization of the data was done separately for train and test sets—this was done to avoid 'leakage' of information from train to test sets.

Reviewer 2. Thank you! If you want to anonymously suggest an additional reference, we'd be happy to add that :-).

Reviewer 3

1. **Empirical evaluation.** We have compared our method on both synthetic and real-world datasets, as well as a less-known setting involving noisy data. These experiments cover most common tail dependencies.

2. **Comparison to Vines.** We do not see ACNet as a competing method to Vines, but rather, a complement which augments the space of bivariate distributions which are available. ACNet, however, is an alternative towards manually selecting bivariate Archimedean Copula, which we demonstrate in our experiments. Furthermore, we believe that the topic of Vines, while interesting, is worth consideration separate from ACNet.

3. **Computational cost.** See point 2 in our response to Reviewer 1.

Reviewer 4

1. **Evaluating ACNet on higher dimensions.** Bivariate distributions are the cornerstone of Copula research. For example, they are the building blocks for Vines, which are a common way of scaling up Copulas. Furthermore, since Archimedean Copulas are symmetric, learning the generator for a 2-dimensional distribution would implicitly learn the generator for a n-dimensional one, assuming the assumption of symmetry holds. It is possible to evaluate ACNet on dimensions $d > 2$ (see point 2 in our response to Reviewer 1) and we will include it in the final paper.

2. **Semiparametric models.** Semi-parameteric approaches often suffer from various limitations such as slow computation and are fundamentally quite different from the neural models we consider. Nonetheless, we agree with the author and will include them in the discussion on background work.

3. **Quantitative evaluation and metrics.** As mentioned in Section 4, we have reported the testing log-likelihood lost in our experiments. This is the most common metric in the machine learning community. The qualitative evaluation is meant to supplement the quantitative evaluation. The reviewer brings up a fair point about other useful metrics in evaluating copula, such as Kendall's Tau and upper/lower tail dependencies (see Appendix 7.5). However these quantities are either difficult to compute (except in very specific parametric models) or focus on goodness of fit of tails as opposed to the whole joint distribution, which is what ACNet is designed for. Nonetheless, we will include a more thorough evaluation in the final version.

4. **Two-phased nature.** The phrase 'two-phased' was used to describe phenomena in normalized data where there appears to be distinct phases in the joint distribution—before and after 0.5, where the degree of positive dependence increases and decreases within each phase. We will make this clear in the final version of the paper.