Dear Prof. Catherine Hurley,
Dear Associate Editor,
Dear Reviewers,

Thank you very much for your constructive and helpful feedback.

In reaction to your comments, we significantly expanded our package and further increased its utility to the R and dependence modeling communities. This covers the following fundamental changes:

1. five new Goodness-of-Fit tests;
2. eight new copula models;
3. the new opportunity to test for rotated copulae in the bivariate case;
4. improved internal architecture of the package by focussing on the function `gof()`, which optimizes the usability of the package for practitioners;
5. a roxygen2 documentation.

Further details are described in the point-by-point responses below. For your convenience, we highlighted changes in the paper in red. We uploaded the new version 0.4-1 of our package to CRAN due to the extended changes, replacing the old 0.3-3. It is worth mentioning that the first version of the package was uploaded to CRAN already in 2015. Therefore, it was polished and debugged, thanks to the extended feedback and bug reports from our users over the years. We tried our best to implement all the changes and provide a smooth working package version in the relatively short time we had for preparing the revision. Unfortunately, we cannot guarantee the absence of bugs. Thus, we kindly appreciate any errors, typos or unclarities you might detect during your review. This package is our little “child” and we are going to work on the current version over the upcoming months and years to further polish, improve and extend it.

Yours sincerely,

Ostap Okhrin, Simon Trimborn, and Martin Waltz

In the following are the responses regarding document 1-review-3.

Article

Comment 1. For the test based on the empirical copula process, the authors consider the Cramer von Mises version while for the test based on Kendall’s process, the authors consider both the Cramer von Mises version and the Kolmogorov-Smirnov version. To justify this, the authors claim that “some of them show clearly worse performance in testing than others”. I agree that in most cases, the test statistic T_n^E shows a worse performance than S_n^E (even if there are a few situations where this is the contrary (Genest et al., 2009). But the test statistic T_n^K also shows inferior performance to S_n^K while the authors have still implemented both. Is there a reason for such a discrepancy?

Reply: *We want to thank you for addressing this issue. We decided to implement the Kolmogorov-Smirnov test based on the empirical copula process as well. Although being in certain situations less powerful, this test statistic definitely should not be missing, as it is for users to decide whether to use it or not. Complementing this addition, we also included four extra tests specifically for Archimedean copulae to increase our package’s comprehensiveness.*

Comment 2. p.9, the authors define the meaning of "hybrid(1,2)" in the middle of the paragraph after the beginning of section "gofCOP class". I think this is not the optimal position for the reader, as the hybrid tests are used quite extensively in the paper. Furthermore, this notation already appears in the output at the top of page 9 and I would prefer to define it there.

Reply: *Thank you for bringing this to our attention. We moved the definition of "hybrid(1,2)" to the desired place (end of page 10 in the current version of the manuscript) to improve this section's readability.*

Comment 3. In Figures 1, 3, 7-10, the authors used extensively "pirateplots". I was not aware of the existence of such plots and I believe that this is the case of most statisticians, but they definitely look very interesting. Could the authors add a paragraph (for example on p.9 where they introduce this) to detail shortly what is a pirate plot and how the reader should interpret such displays, so that the article is more self-contained? Such background would be very useful.

Reply: *Thank you, "pirateplots" indeed look nice and provide a lot of important information. In reference to the referees' following fourth comment, we added a section to introduce pirateplots when they first appear in the paper (page 11 in the new manuscript). The details of page 14 in the old manuscript are merged into this section, making sure the meaning of the illustrations is clear to the reader.*

Comment 4. p.9, the authors gives an example of the pirateplots but only detail what it is in page 14. I think the authors could merge the two discussions.

Reply: *Agreed and changed accordingly.*

Comment 5. If I understand well, it seems to me that the "hyb1" does not refer to hybrid tests but contains the original test p-values, is it true? I think the authors could add one or two sentences to make it more explicit. Maybe it would be easier to the reader if "hyb1" was replaced by something more explicit, such as "singleTests".

Reply: *Your understanding is correct, "hyb1" indeed references the p-values of the original single tests, being interpreted as a hybrid test of size 1. We followed your suggestion and clarified the naming to "singleTests" in the corresponding column header of the pirateplots. Additionally, we added a sentence to explain the specification of "hybrid = c(1, 3, 5)" in detail.*

Comment 6. I think that the function gofCheckTime is a good idea and will be very useful in practice, but I wonder about its computation time itself. Is it long to have such an estimate of the computation time?

Reply: *The computation of the function "gofCheckTime" is usually relatively fast, taking only a couple of seconds to give a reasonable estimate of a given testing task. However, specifically the computation of the "gofPIOSTn" test is more demanding, thus calling "gofCheckTime" can take in this case – depending on the engine, data, and copula – a bit longer in comparison to the remaining tests. You are right that this would be of concern in practice. To make the user aware of this issue, we added now additional warnings in "gofCheckTime" as well as directly in "gofPIOSTn". Thank you for bringing this to our attention. Furthermore, we would like to mention that the change in the internal architecture of* *the package lead to a more efficient computation flow. Consequently, the overall computation times slightly decreased in comparison to the 0.3-3 version.*

Minor Points

Comment 1. p.6, v is not defined in Equation (6).

Reply: *Thank you, we changed it accordingly.*

Comment 2. p.11, line 17, I think you could mention the types too: x should be a matrix and copula a character string.

Reply: *Agreed and changed accordingly.*

Package

Comment 1. The user documentation of the package is very good, but I found the internal documentation of the code quite lacking. This may be linked to the lack of roxygen2 documentation.

Reply: *Thank you for this great suggestion. We thoroughly improved the code's internal documentation and decided to use roxygen2 documentation to increase the internal comprehensibility of our package.*

Comment 2. The very long lines of code (sometimes nearly 2000 characters on the same line!) makes it also difficult to read, while most style guides recommend a limit of 80 or 100 characters to be readable on one screen without needing horizontal scrolling.

Reply: *Thanks for pointing this out to us. We divided these extensive code lines into several sublines to increase the readability of these parts of our code so that there is no need to scroll horizontally.*

In the following are the responses regarding document 1-review-4.

Comment: I first note that the goodness-of-fit methodologies that are implemented in the package are accurately described. From my point-of-view, the code is well written, easy to use and the outputs are clear. An investigation of the computing time as a function of bootstrap parameters is done. The hybrid method is a valuable methodology and the graphical representation of the tests' results are interesting and useful. The examples are helpful as well.

However, I do not completely agree that the code offers a « broad range of tests » (see the Conclusion), hence I am mitigated about the importance to R the community. In fact, for the choice of the available GoF tests, (a) several ones were already implemented, (b) other ones are mainly taken from Zhang et al. (2016) and (c) many are based on the Rosenblatt transform, the latter performing generally poorly. Also, the number of testable copula families is restricted to the Normal, Student, Frank, Clayton and Gumbel, which is unfortunate in view of the large spectrum of available models.

Reply: *Thank you for this detailed feedback. Our original intention with the package was to provide only tests with the general-purpose, not restricted to specific copula families. But after your comment, we agree that our original idea strongly limits the flexibility of the package. Therefore, we drastically* *expanded the scope of our package to increase its overall comprehensibility. In detail, we added five tests that, up to the best of our knowledge, were not implemented in the R package landscape beforehand. This includes the test based on the Kolmogorov-Smirnov statistic of the empirical copula process and four tests based on the approach considered in Hering & Hofert (2015). Furthermore, we added eight further copula models to the package: Joe, AMH, Galambos, Husler-Reiss, Tawn, t-EV, FGM, and Plackett. Thus 13 copula models are currently available, covering the most essential elliptical, Archimedean, and extreme-value copulae. Furthermore, we implemented the option to use each copula in rotated form by either 90, 180, or 270 degrees in the bivariate case. Finally, as not every test can be used for every copula, we provide a user-friendly overview of the possible test-copula combinations, including their available dimensions through the function CopulaTestTable(). We would also like to remark that the user always has the opportunity to make use of the whole testing pipeline by specifying a customized test statistic and using the function gofCustomTest().*

Comment (i) Page 2. The way it is mentioned seems to discredit the tests by Hering & Hofert (2015) (intended to d-dimensional Archimedean copulas), Quessy et al. (2019) (specific to Chi-square copulas) and Jaser et al. (2017) (for elliptical copulas). Please rewrite and note in passing that the test by Jaser et al. (2017) (see its rank version by Quessy (2020)) is not a GoF test for a specific copula family, but rather for a *shape*.

Reply: *Discrediting any of the authors mentioned above is certainly not our intention, and we want to thank you for pointing this out. Let us stress here once more that the original purpose of the package was to provide general-purpose tests. However, following the thorough expansion of our package and our decision to include the approach of Hering & Hofert (2015), this paragraph is generally not suitable for the paper anymore, and therefore we deleted it entirely. Our decision to include tests based on Hering & Hofert (2015) is motivated by the circumstance that the considered transformation is easily implementable while allowing for a variety of different tests, similar to the Rosenblatt-based approaches. Thus, we can provide these additional tests for each of the five Archimedean copulae to our users.*

Comment (ii) I would use something like $\{C_\theta; \theta \in A \subset R^p\}$ instead of C_0 for the copula family under the null hypothesis.

Reply: *Agreed. We specified C_0 accordingly.*

Comment (iii) Page 4. Please define PIOST.

Reply: *We defined more generally PIOS as "pseudo in-and-out-of-sample", so that the namings of PIOSTn and PIOSRn referencing the corresponding test statistics are transparent to the reader.*

Comment (iv) Page 4. The sentence « Few of the tests use slightly different H_0-hypotheses » sounds quite enigmatic this place in the text.

Reply: *Agreed, we removed this sentence and referred to H_0 in the given paragraph as the "general" H_0 hypothesis. This is intended to emphasize the non-specific character of this introductory statement, which is detailed for the respective tests in the following.*

Comment (v) What is commonly referred to the empirical copula process and Kendall's process are defined with a fixed θ (and note its estimator).

Reply: *Thanks for this note, we changed the paragraphs on the empirical copula process and Kendall's process accordingly.*

Comment (vi) Page 4, after (6). «… (not ~~necessarily~~ uniform in general) … »

Reply: *Thank you, we changed it accordingly.*

Comment (vii) Just before the section on the Rosenblatt transform. Please give a reference that justifies the equivalence between H_0 and H_0".

Reply: *Thank you, we would like to refer here to Genest et al. (2009), who make this statement. We added the reference at the corresponding point in the manuscript as well.*

Comment (viii) Why do you take the absolute values in Step 4 of the parametric bootstrap?

Reply: *We use absolute values since it allows us to handle the two-sided case, in which an H_0 hypothesis might be rejected due to very large or very small values of the test statistics. Such an approach is pursued in the original paper proposing PIOS tests, Zhang et al. (2016).*

Comment (ix) Page 7, after the steps of the parametric bootstrap. Maybe give a short explanation about « double bootstrapping » for readers less familiar with the method.

Reply: *Thanks for this suggestion. We added an explanation, please see page 8.*

Comment (x) The illustration with the data in Figure 2 should be done with the ranks, even if the marginals are uniform on [0,1].

Reply: *Thanks for pointing this out. We modified Figure 2 to show the data directly rank-based and additionally with standard normal margins.*

Comment (xi) Unlike what is claimed at some places, it is not the fact that the data « are not in the domain [0,1] » that justifies the use of ranks, but rather because the marginal distributions are left completely unknown. Conversely, data could be inside [0,1] but not uniform.

Reply: *Thank you for bringing this to our attention. The prior version indeed lacked preciseness in this regard. However, we decided to use the non-parametric margin transformation per default to pursue a consistent approach through all examples in the paper.*

Comment (xii) Page 20. You probably mean that « the Frank copula is radially symmetric, as is the case with elliptical copulas. »

Reply: *Yes, indeed, the sentence is adjusted.*

REFERENCES

GENEST, C., RÉMILLARD, B. & BEAUDOIN, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics* **44**, 199-213.

HERING, C. & HOFERT, M. (2015). Goodness-of-fit tests for archimedean copulas in high dimensions. In K. Glau, M. Scherer, and R. Zagst (editors), *Innovations in Quantitative Risk Management*, pp. 357–373, Cham. Springer International Publishing.

JASER, M., HAUG, S. & MIN, A. (2017). A simple non-parametric goodness-of-fit test for elliptical copulas. *Dependence Modeling* **5**, 330-353.

QUESSY, J.-F. (2020). On nonparametric tests of multivariate meta-ellipticity. *Statistical Papers*, 1-28.

QUESSY, J.-F., RIVEST, L.-P. & TOUPIN, M.-H. (2019). Goodness-of-fit tests for the family of multivariate chi-square copulas. *Computational Statistics & Data Analysis* **140**, 21-40.

ZHANG, S., OKHRIN, O., ZHOU, Q. M. & SONG, P. X.-K. (2016). Goodness-of-fit test for specification of semiparametric copula dependence models. *Journal of Econometrics* **193**, 215-233.