

Here are our responses to the comments from the Associate Editor:

AE's general comment: I have heard from two referees. The views are mixed. Referee 2 is not enthusiastic about the paper and recommends rejection while referee 1 recommends a major revision. I also read the paper carefully and understands both referees concerns

Our response: Thank you for reading our paper so carefully. We are very grateful for your constructive comments which has lead to a much improved paper.

Referee 2 cover letter: Definitely the topic is very interesting. However, I do not see a clear contribution of this paper to the literature on estimating dependence and testing independence.

I don't see the added value of this paper when is compared to the nonparametric approaches that use Kernel functions (e.g. Normal Kernel) to locally approximate joint distribution functions.

In addition we have many available tests for conditional independence between a vector and scalar variables (and also between two vectors of random variables) conditional on a third vector random variable. However, the test proposed in the paper is limited to testing conditional independence between two scalars only. In addition, the simulation results are definitely not better than the existing tests.

Also, I believe the asymptotic and bootstrap theories are not complete. I also made other comments.

For the above reasons, I believe that the contributions in this paper is definitely below the level of JBES.

Our response: We have made a response directly to the 6 points in the report of referee 2. We briefly summarize this below, at the same time responding to the cover letter of referee 2:

1. **The novelty of the paper.** We look as the main contribution of our paper the introduction of the new concept of local partial Gaussian correlation (LGPC). With this concept we have sought to make a nonlinear and local extension of the ordinary (global) partial correlation, whose use in measuring and describing dependence is well known for Gaussian distributions. We claim that it is fundamentally different both in concept and in use from the existing tests of conditional independence. It is hard to obtain a measure of conditional dependence that can be easily interpreted from these test as they would have to be based on distance measures that for example cannot distinguish between negative and positive conditional dependence.

Using the kernel estimates for the purpose of measuring conditional dependence is not easy in the same way as it is not easy to use the joint distribution function for a numeric measure of dependence. See item 1 in our response to referee 2.

Clearly we did not get our message concerning our main contribution through to this referee. In retrospect we see that this may be our own fault. We did not succeed in adequately presenting the main purpose of our paper. Our main message is not that of proposing just another test of conditional independence, and we have tried to make this more clear in the present version. Our test did do well in a competition with other test, and it may be possible to have it improved with better focusing and use of more local parameters, but, as perhaps over-stated, the main contribution, we believe, is the introduction and use of the new concept of LGPC in measuring the conditional dependence by considering a nonlinear local version of the partial correlation, a measure that reduces to the ordinary partial correlation for Gaussian variables.

2. **Estimation and testing in the full vector case:** We have developed a theory for this. More details can be found in the the response to item 3 in the report of referee 2. We have also added a new example where X_1 and X_2 are not both scalars.
3. **Asymptotic theory and bootstrap validity.** This has been developed in Sections E, F and H in the Supplement, and details can be found in our response to item 4 in the report of referee 2. Actually, we are happy that referee 2 raised this point. We believe that it has strengthened the paper considerably. We add that we have also examined local power and the Pitman criterion in Section G of the supplement. See also our response to item 5 of referee 2.

AE further general comment: Referee 1 enjoys reading the paper and have some comments that can be relatively easily addressed.

Given the fact that both referees think that the paper is interesting and I also feel it is interesting, I lean toward the positive referee and do not object to seeing a revision of the paper. Nevertheless, the authors should keep in mind that they must address the referees' major concerns to their satisfaction in order for the paper to pass the second round of review.

My main comments are as follows, some of which overlap those of the referees.

Our response: We are grateful for the opportunity to revise, and we have responded to all of the comments of the two referees, and we will respond to your comments below.

Comment 1: I like the idea to extend the local likelihood framework of Hjort and Jones (1996) to estimate the local correlation and to extend the measure of local Gaussian correlation of Tjøstheim and Hufthammer (2013) to the measure of local Gaussian partial correlation (LGPC).

Our response: Thank you.

Comment 2: As both referees mention (see R1's point 1 and R2's point 3), one limitation of the proposed testing procedure is that it only tests the conditional independence between two scalar variables (X_1, X_2) given a random vector (X_1, \dots, X_p) . It is important to generalize the test to allow both X_1 and X_2 to be random vectors.

Our response: Thank you for this important input. We now introduce and define the LGPC (matrix) for a general \mathbf{X}_1 , \mathbf{X}_2 and \mathbf{X}_3 of dimension d_1 , d_2 and d_3 in Section 2.1. The full multivariate framework is analyzed in more detail in Section A: “The multivariate LGPC” in the Supplement. Various special cases are discussed in Section A.2 of that Supplement. There are really no new concepts involved, and large parts of the ensuing estimation and testing theory in the Supplement are carried out in this framework.

In particular the asymptotic theory for a conditional independence test where both \mathbf{X}_1 and \mathbf{X}_2 (and \mathbf{X}_3) are vectors is given in Section F of the Supplementary material (Theorem F.1 and Theorem F.2). The asymptotic local power is discussed in Section G, and the validity of the bootstrap for the testing statistic are given at this level of generality (Theorems G.1, H.1 and H.2). However, the notation becomes more complex and expressions less explicit, and for these reasons and for reasons of space much of the material in the main article is stated for the case of a scalar X_1 and X_2 . There is a new example of Granger causality in Section 4, though, where the assumption of a scalar X_1 and X_2 is dropped. The curse of dimensionality can be avoided at the cost of a simplified pairwise model introduced described and used in the main article and in the supplement. Throughout the main paper the reader is referred to appropriate sections of the Supplement for the general vector case.

Comment 3: p.5: As Referee 1 points out, the density in (6) is not the real density of \mathbf{Z} . It is a local Gaussian approximation.

Our response: You are certainly right. This was not well written. We have changed “representation” to “approximation” and also linked this directly to the defining approximative equation (4).

Comment 4. p.10: Since you claim that \mathbf{b} is a diagonal matrix of bandwidths (not squared bandwidths). It seems that you want to define $K_{\mathbf{b}} = |\mathbf{b}|^{-1}k(\mathbf{b}^{-1}\mathbf{x})$.

Our response: Yes, corrections have been implemented. Thank you

Comment 5: pp.12-14:

- (a) p.13: It seems that both $\mathbf{J}_{\mathbf{b}}$ and $\mathbf{M}_{\mathbf{b}}$ are of exact order $O(1)$ (so that the second term in the definition of $\mathbf{M}_{\mathbf{b}}$ is asymptotically negligible). This suggests that the estimator of the local correlations have convergence rate $(nb^5)^{-1/2}$ under certain smoothing bias conditions. Why is the convergence rate here much slower than the usual $(nb^3)^{-1/2}$ -rate for the kernel estimator for a trivariate density? In addition, it is possible to estimate the local conditional correlation or not? Note that for nonparametric tests based on the conditional distribution, typically only the dimension of the conditioning variable matters for the convergence rate.
- (b) p.14, line 2: $1/(nb^3)$ should be $1/\sqrt{nb^3}$, which is the convergence rate of a trivariate density..

- (c) p.14, line 4: $1/(nb^5)$ should be $1/\sqrt{nb^5}$.
- (d) The results are too limited as it is only for the case with $p = 3$. A general result for the generic p is needed.
- (e) Theorem 3.1 is a pointwise result. For the inference purpose, one typically needs a uniform result.

Our response:

- (a) The reason for the slower convergence can basically be found in the expressions for \mathbf{J}_b and \mathbf{M}_b . These expressions contain the matrix $\mathbf{u}(\cdot)\mathbf{u}^T(\cdot)$. This matrix has rank 1, and as $b \rightarrow 0$ this implies that the matrices \mathbf{J}_b and \mathbf{M}_b are singular and of order $O(b^2)$. It is therefore necessary to do a Taylor expansion to find the correct scaling factor, and this in turn leads to the stated convergence rate of $(nb^5)^{-1/2}$. (Note that there was a typo in eq. (20). It has now been corrected.) This is explained in the proof of Theorem 3.1 and in more detail in Section 4 of Tjøstheim and Hufthammer (2013) in a slightly different situation. Similar results and with rather different derivations can also be found in Jones (1996, 1998). In the pairwise situation referred to after the statement of Theorem 3.1, the matrices \mathbf{J}_b and \mathbf{M}_b collapse to scalars, the singularity disappears and one is back to the usual rates for the case in nonparametric kernel estimation with the dimension of the conditioning variable determining the convergence rate.

It is certainly possible to estimate the local conditional correlation (and in the general vector case) using local likelihood for local correlations and then transforming to the LGPC. In the general vector Case this is explained in Sections A.2 and A.3 in the Supplement.

- (b) This has been corrected. Thank you.
- (c) This has been corrected. Thank you.
- (d) Agree and the general vector case has now been implemented. We refer to our response to your comment 2 above.
- (e) We agree that our results are pointwise results, thus resulting in pointwise confidence intervals for example. We have to leave uniform results and confidence bands to future investigations.

Comment 6: p.16: Is there a data-driven choice of bandwidth? The choice of $c = 4$ seems too arbitrary.

Our response: There is a data-driven choice for bandwidth in a density estimation context in Otneim and Tjøstheim (2017). It is based on leave-one-out cross-validation. This is referred to in the beginning of Section 4. It remains to establish such a procedure for the present situation. In the paper we have typically used the asymptotic rates $b = cn^{-1/9}$ for the full trivariate fit and $b = cn^{-1/6}$ for

the bivariate simplification. In the testing results (Table 2, 2* and 3* we have used $c = 1.0$ and $c = 1.4$ to examine sensitivity to bandwidth selection. They gave similar results. Note also that for some of the other competing test several choices of tuning parameters were used. For the plots of the estimated LGPC in Section 4 it is true that we used more smoothing to obtain cleaner plots. Thus $c = 4$ for Figure 1, but a smaller bandwidth are used in testing and as mentioned in the description of that example the hypothesis of conditional independence is not rejected for any reasonable significance level.

Comment 7: p.24-26: The authors propose a new test for conditional independence based on the LGPC. The test statistics should be formally studied under the null hypothesis of conditional independence and the usual Pitman sequence of local alternatives. For this purpose you need uniform consistent estimate of the local correlations. As referee 2 mentions in his/her fourth comment, you also need to justify the asymptotic validity of the bootstrap.

Our response: The asymptotic properties of the test statistic has been formally studied in Section F of the Supplement. The validity of the bootstrap for the test statistic is established in Section H of the Supplement (for estimation in Section E of the Supplement. The convergence rate of the test statistic is given in Section 5.2 in the main article. It is $n^{-1/2}$ for the pairwise testing case and $(nb^2)^{-1/2}$ for the trivariate testing case. This is briefly compared to rates for other test statistics. In Section G we look at local asymptotic power and the Pitman criterion. We essentially use the approach of Wang and Hong (2018). (They avoid the uniformity argument presumably by using the weaker condition D' in Noether, G.E. (1955).)

Comment 8: p.26: For the comparison with existing tests, you need to compare asymptotic power properties too.

Our response: We refer to our response to comment 7, where we have compared convergence rates and indirectly local power

Comment 9: The paper contains a lot of typos. Referee 1 pointed out some of them and the authors should correct others too.

Our response: We have corrected the typos kindly pointed out to us by Referee 1. We have read carefully through the revised version, including the Supplement in order to eliminate as many typos as possible.

References

- Jones, M. (1996). The local dependence function. *Biometrika*, **83**, 899-904.
 Jones, M. (1998). Constant local dependence. *Journal of Multivariate Analysis*, **64**, 148-155.
 Noether, G.E. (1955). On a theorem of Pitman. *Annals of Mathematical Statistics*, **26**, 64-68.

Otneim, H. and Tjøstheim, D. (2017). The locally Gaussian density estimator for multivariate data. *Statistics and Computing*, **27**, 1595-1616.

Tjøstheim, D. and Hufthammer, K.O. (2013) Local Gaussian correlation: A new measure of dependence. *Journal of Econometrics*, **172**, 33-48.

Wang, X. and Hong, Y. (2018). Characteristic function based testing for conditional independence: A nonparametric regression approach. *Econometric Theory*, **34**, 815-849.