Here are our responses to the comments from the referees:

Responses to referee 1

Referee's general comment: The authors propose a new conditional dependence measure (called the local Gaussian partial coefficient, LGPC). They then construct an alternative test for conditional independence and test Granger causality in time series. I enjoy reading the paper and only have a few specific comments/suggestions that matter for the exposition of the paper. And I recommend a revision. In the following, I list the expositional issues that I would like to point out and discuss with the authors.

> **Our response:** We are extremely grateful for your thoughtful comments and for your careful reading of our manuscript. Your comments have lead to very substantial improvements in the paper. Below is our response to the detailed comments 1 – 12.

Comment 1: The whole paper discussed the LGPC coefficient as the local version of the partial correlation coefficient between X_1 and X_2 given $\mathbf{X}^{(2)}$, with $\mathbf{X}^{(1)} = (X_1, X_2)$ and $\mathbf{X}^{(2)} = (X_3, \ldots, X_p)$. While it is easy to understand, it seems that the discussed dependence measure is only limited to bivariate case conditional on the third vector. Many of the existing conditional independence tests mentioned in Section 4 are, however, designed to work for general dimensions; for example, when testing $\mathbf{X} \perp \mathbf{Y}|\mathbf{Z}$ with $(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \in \mathbb{R}^{d_x+d_y+d_z}$. Of course, "curse of dimensionality'' would always be a serious issue depending on the specific testing methodology adopted; among them the empirical process based tests would be less affected by the dimension of $\mathbf{Z}$ in theory. Perhaps, more importantly, the aforementioned conditional independence tests do not place restrictions on the dimension of $\mathbf{X}$ and $\mathbf{Y}$, and theoretically speaking, they should work for very general settings in which the components X_1 and X_2 in $\mathbf{X}^{(1)}$ can be multivariate. How can the current LGPC measure adapt to the multivariate X_1 and X_2 case? Could you provide some insights or remarks towards this direction?

> **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 LPGC" 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 bath $\mathbf{X}_1$ and $\mathbf{X}_2$ (and $\mathbf{X}_3$ are vectors, is given in Section F of the Supplement (Theorem F.1 and Theorem F.2). The asymptotic 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

1

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 we refer to the appropriate sections of the Supplement for the general vector case.

Comment 2: In Equation (6), isn't it an approximation of the density f_Z of $\mathbf{Z}$ in the neighbourhood of $\mathbf{z}$? The way it writes now seems to indicate that the expression is exactly the density $f_{\mathbf{Z}}$ of $\mathbf{Z}$, but $f_{\mathbf{Z}}(\mathbf{z})$ can certainly differ from $\psi(\mathbf{z}, R(\mathbf{z}))$ in the general case.

> **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 3: In the definition of $F_{j,n}(x)$ of page 5, X_i should be X_{ji}. Also the empirical analogue for each Z_j can only be written as $Z_{ij,n} = \Phi^{-1}(F_{j,n}(X_{ji}))$ for $j = 1, \ldots, p$, and $i = 1, \ldots, n$.

> **Our response:** Thank you. This has been corrected on page 6 of the new version.

Comment 4: Above Equation (10), for the conditional density, use notation $f_{Z_1|\mathbf{Z}_3}(z_1|\mathbf{z}_3)$ to be consistent with subsequent notations.

> **Our response:** Has been done. Thank you.

Comment 5: In Equation (15), change $\hat{\mathbf{z}}$ to $\mathbf{z}$. And isn't the definition of kernel function $K_b(\mathbf{x} = |\mathbf{b}|^{-1}K(\mathbf{b}^{-1}\mathbf{x})$?

> **Our response:** Yes, both corrections have been implemented.

Comment 6: In Equation (17), change $\hat{\mathbf{z}}$ to $\mathbf{z}$, $(\hat{z}_j, \hat{z}_k)$ to (z_j, z_k), and there is a ψ missing in log.

> **Our response:** We have corrected these mistakes. Thanks.

Comment 7: In Equation (19), I suspect the expression should be $\mathbf{M}_b^{-1/2}\mathbf{J}_b$ instead of $\mathbf{J}_b M_b^{-1/2}$. Pls double check. Although the exact expression of the asymptotic covariance matrix $\mathbf{J}_b^{-1}\mathbf{M}_b\mathbf{J}_b^{-1}$ is not available, could you state the orders of $\mathbf{M}_b$ and $\mathbf{J}_b$? It is hard to understand the convergence rate without knowing the respective orders of $\mathbf{M}_b$ and $\mathbf{J}_b$?

> **Our response:** Thanks a lot for spotting this error in the sequencing of those two matrices. (I am afraid that this mistake has gone undetected in some earlier publications). The orders of $\mathbf{M}_b$ and $\mathbf{J}_b$ are $O(b^2)$. This has now been stated with a reference to the Supplement for their derivation. We have also corrected a typo in eq. (20).

Comment 8: In page 18, line 10 from below, starting from "It is straightforward to see that". The sentence is very confusing as it holds only when $\rho = 0$ in the context of Figure 3.

> **Our response:** We believe that it holds for every ρ via a conditioning argument. We have rephrased the sentence.

Comment 9: In page 25, line 13 from below, starting from "Departures from many types of conditional independence". This sentence is also confusing as conditional independence is a restriction to the underlying populations under the null hypothesis. Are there many other different types of conditional independence?

> **Our response:** We are sorry for this absolutely silly statement. It has been corrected.

Comment 10: In Table 1, correct the typos in DGPs 3(change X_{t-1}^2 and 10 (Change Y_t^2).

> **Our response:** Thank you: Done.

Comment 11: In Table 2, what are the tests SCM and SKS?. I couldn't find these definitions anywhere.

> **Our response:** Sorry, has now been included.

Comment 12: In the proof of Theorem 3.2, to show that

$$\frac{1}{n}\sum_{i=1}^n (F_n(X_i) - F(x_i)) \to_p 0,$$

a conventional U-statistic argument (e.g. Hoeffding decomposition) for dependent data (e.g. strong mixing) is perhaps much easier and preferable, by noting that

$$\frac{1}{n}\sum_{i=1}^n (F_n(X_i) - F(X_i)) = \frac{1}{n^2}\sum_{i=1}^n\sum_{j\neq i}^n \{I(X_j \leq X_i) - F(X_i)\} + \frac{1}{n^2}\sum_{i=1}^n\{1 - F(x_i)\}$$

$$= \frac{1}{n}\sum\{0.5 - F(X_i)\} + O_p(n^{-1})$$

$$= O_p(n^{-1/2}) + O_p(n^{-1}) = O_p(n^{-1/2}).$$

This also implies that the order of $\frac{1}{n}\sum_{i=1}^n (F_n(X_i) - F(X_i))$ is sharper than the obtained order $n^{-1/2}(2\log\log n)^{1/2}$.

> **Our response:** Thanks very much indeed. This is much easier, more elegant and preferable. We have taken the liberty to include it with acknowledgement to "an anonymous referee".

Responses to referee 2

Referee's general comment: The paper is interesting and well written. However, I have the following comments:

> **Our response:** Thank you! We are grateful for your constructive comments. They have lead to great improvements of the paper.

Comment 1: What is the difference between your approach and existing nonparametric approaches that use kernel functions (e.g. Normal Kernel) to locally approximate joint distribution functions?

> **Our response:** These differences have been looked at in two papers. For the estimation of the joint distributions we refer to Otneim and Tj{ø}stheim (2017), and for the conditional distribution to Otneim and Tj{ø}stheim (2018). The authors of these two papers show that the local Gaussian approach is much less vulnerable to the curse of dimensionality using a pairwise local Gaussian approach. However, both the kernel and local Gaussian approach to estimating conditional distributions are ill-fitted to actually *measuring* the conditional dependence numerically, for the same reason that the joint distribution is not a well fitted tool to actually measuring the dependence between two or more stochastic variables numerically.

Comment 2: Using estimated Z (obtained from the empirical distribution of X instead of the true Z introduces some estimation effect/error. Might be this error is not negligible for finite samples. Did you examine the impact of this estimation error on the estimation of dependence on testing independence when the sample size is small?

> **Our response:** Asymptotically it is negligible as demonstrated by Theorem 3.2 and its proof. We can examine finite sample effects through our simulation experiments where the truth is known. This is the case for the estimation of the level (5% is correct) for the DGPs created under the conditional independence of the null hypothesis. The results from these experiments are shown in Table 2 in the main article and in Tables $2^* - 4^*$ in the Supplement, Section K. These experiments were carried out for sample sizes of 100, 200 and 400, and even for a sample size as small as 100, results close to the correct ones were obtained.

Comment 3: One limitation of your testing procedure is you can only test independence between two scalars (X1 and X2). This is a big issue for testing Granger non-causality, because $X1$ and $X2$ has to be a vector if many lags are needed for Granger causality analysis.

> **Our response:** Thank you for this important comment to which we agree. As a consequence we have rewritten the paper extensively. 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 LPGC'' of 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 Supplement (Theorem F.1 and Theorem F.2). The local asymptotic 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 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 paper the reader is referred to the appropriate sections of the Supplement for the general vector case.

Comment 4: The asymptotic and bootstrap theories for the test statistic of conditional independence are missing. I appreciate the authors refer to some papers for this, but it will be valuable for readers if these theories can be adapted in the context of the current paper and the results can be included in the paper.

> **Our response:** We have followed up your suggestion which we think has resulted in a great improvement of the paper. For reasons of space we have had to put these developments in the Supplement, mostly for a general $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3\}$, but with appropriate references in the main article to where they can be found. The validity of the bootstrap for estimation can be found in Section E of that Supplement, the asymptotic theory for the test statistic in Section F, the bootstrap validity for the test statistic in Section G.

Comment 5: A theoretical discussion on how your tests compares to existing test (in terms of the rate of convergence . . .) will be appreciated.

> **Our response:** Obviously this is another important point. As a consequence we have included a discussion of this in Section 5.2 in the main article and in Section G of the Supplement on local asymptotic power and the Pitman criterion.

Comment 6: In the simulation part, I think you should consider other distributions other than the normal distribution. Your test is optimal when the joint distribution is normally distributed, Thus, I believe it is an unfair comparison with Su and White tests and other tests that are fully non-parametric.

> **Our response:** It is somewhat hard not to disagree with this comment. The optimal test in the Gaussian case is the LIN, the ordinary linear Granger test. It does well (best, see Table 2 of main article) on DGP 5 which is linear Gaussian, but it does miserably on the other DGPs, which have a non-Gaussian distribution. Our test does slightly worse than the Granger test on DGP 5, but it does well on

the other DGPs which has a non-Gaussian joint distribution. We think that the competition with Sun and White is fair, when we have chosen *their* examples to compete against.

References

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

Otneim, H. and Tjøstheim, D. (2018). Conditional density estimation using local Gaussian correlation. Statistics and Computing, **28**, 303-321.