Response to reviewer comments to the manuscript *2019-146: lg: An R package for Local Gaussian Approximations* submitted to *The R Journal*

I am very grateful to the Editor of *The R Journal* for considering this paper for publication, and to the reviewer for his/her careful reading and several useful comments. I have revised the paper according to the comments, as detailed in a point-by-point fashion below.

> *Summary*
>
> *The aim of this paper to introduce and to illustrate the use of the lg package for R which implements the computation and visualization of the concept of 'local Gaussian correlation'. The local Gaussian correlation (LGC) is a measure of linear and non-linear dependence, proposed recently by Tjøstheim and Hufthammer (2013). Specifically, the paper presents how the GLC can be estimated from empirical or simulated data and how the 'dependence structure' described by the GLC can be visualized. Further, the application of statistical tests for the significance of correlation (i.e. GLC) is demonstrated.*

My response: This is a good summary that captures the essence of the paper.

> *Overall evaluation*
>
> *Measuring and visualizing the structure(s) of temporal dependencies and dependencies among variables – or aspects thereof – is highly relevant in all areas and sub-disciplines of statistics (e.g. biometry, econometrics etc.). Therefore, providing a software package that implements the local Gaussian correlation, a measure of linear and non-linear dependencies recently proposed by Tjøstheim and Hufthammer (2013) is very useful since it allows a broader audience to apply this concept in empirical research. In this regard presenting such a software package in a scientific journal seems to further support applied statisticians in using this new concept of dependence. However, the author fails to convince me that the manuscript constitutes a serious scientific piece at this stage which deserves publication in a well respected scientific journal. A number of reasons have lead to this conclusion.*

My response: The reviewer is exactly right. There is a lot of interest in the measurement, visualization and modeling of statistical dependence in many empirical disciplines. The purpose of this paper is to make recent developments within local Gaussian approximations that have appeared in the statistical literature in recent years available to a broader audience of practitioners. I appreciate the comments by the reviewer below, and I hope that the revised paper satisfies the high scientific standards of *The R Journal*.

> *To exemplify this, the section on 'Statistical background' reads as if compiled from different sources; In particular, the paragraphs are not well connected in my view.*

My response: The purpose of this section is to give the reader a brief overview of the statistical methods that have been implemented in the **lg**-package, without going too much in to the technical details. I have revised this section carefully, with particular emphasis on improving the transitions between the paragraphs.

> *Further the explanation of the basic idea of the GLC measure in the section 'Statistical background' could be improved, thereby making the concept accessible for a broader audience and making this paper more 'self-contained'.*

My response: Thank you for this input. The reviewer is correct to point out that the section on "Statistical background" is fairly short and does not go into great detail. The composition of this section is the result of a careful trade-off. On the one hand I fully agree that a more comprehensive section would be desireable, because that would allow the reader to appreciate the technical nuances within the topic of the LGC. On the other hand, it is the purpose of this paper to demonstrate the practical implementation of a large and diverse set of statistical methods. It is my conclusion that this section should not be much longer than it is in its current form (approx. 4 pages), which naturally puts a limit on the level of detail. I have instead opted to provide a rather acessible birds-eye view of the relevant literature, which I believe is useful to the readership

of *The R Journal*. This section has been heavily revised as prompted by the reviewer in his/her comment above..

> *Also, the whole manuscript is not well-structured: The section with the illustration of the lg package is called 'The lg-object' and more importantly is confusing since, for instance, all three different estimation methods implemented are outlined, although not necessary for the (empirical) example chosen by the author;*

My reponse: The paper is structured in the same way as the package it describes. There are two steps to every analysis; in the first step the practitioner must make several modeling choices which to a large degree is guided by the respective references. This step is in common for all applications. I have created a new title for this section to clarify the structure; it is now called "The first step: Creating the lg-object". The various applications are then described separately and organized under informative headlines.

I am not entirely sure how to address the reviewer's second remark. The examples in this paper are chosen to illustrate key steps in the implementation of the methods. It is not my intention to demonstrate all possible combinations of the modeling choices in examples, as that would likely reduce the value of the paper as a general, readable and fairly brief introduction to the relevant R-functions.

> *A section with a summary or concluding remarks is fully missing!*

My response: I have added some conluding remarks in the end of the paper.

> *Furthermore, the code chunks presented in the paper could be augmented by comments explaining what the different lines of code do; For example, in the section 'Graphics' it is not clear what expand.grid does.*

My response: Thank you for pointing this out. Several of the somewhat longer code chunks were not properly commented. Has been fixed.

> *In all, this version of the manuscript, unfortunately, appears to be incomplete and as if submitted under time pressure.*

My response: I can ensure the reviewer and the editor that this is not the case. The manuscript has been prepared carefully, under no time pressure whatsoever. I do thank the reviewer for several useful comments, which I believe has contributed to a significant improvement in the presentation of the material.

Here are my responses to the specific issues addressed by the reviewer:

> *1. As far as I see is your manuscript missing a final section that summarizes the main ideas (e.g. functions here) from the paper and provides some concluding remarks and perhaps an outlook (e.g. what will be added to package in the future). I strongly recommend to provide such a section.*

My response: Has been added.

> *2. Are you sure that Equation (1) is correct? I think x should be x_0 . Further I do not understand y and assume y should be X_i?*

My response: Yes, this equation is correct. The x is the location of the local estimate as defined in the discussion leading up to equation (1), and y is the integration variable in the penalty term of the local likelihood function. This notation is consistent with the literature on local likelihood estimation. The second term does in fact not depend directly on the observations, and the reviewer is quite right to point out that this is not obvious. I have therefore added a comment about this, as well as a reference to a much more detailed justification of the local likelihood function.

> *3. Are you sure that 'the same test statistic and bootstrap procedure as for independence' can be used with time series data? Is not block-bootstrap necessary in this case? Please explain briefly! (p. 3)*

My response: In this particular case we can use the standard bootstrap, because serial dependence in a time series can be investigated by considering the statistical dependence between X_t and X_{t-k} directly. This is a parallell to the autocorrelation function, which for a stationary time series is defined as the ordinary

Pearson correlation $\gamma_k = Corr(X_t, X_{t-k})$. I have changed the formulation here so that it directly states that we obtain critical values by permutations. The block bootstrap is then introduced in the next paragraph, where it is needed to test for serial cross-dependence.

> *4. I understand that the bandwidth controls how many observations are used to compute the correlation coefficient of one grid. Does lg also allow to extract the (exact) number of observations used to compute the different correlations coefficients (as depicted e.g. in Fig. 1)? I feel that from the perspective of an applied statistician, this figure could provide an intuition about how reliable one or the other coefficient is.*

My response: The bandwidth does not define a finite region over which the usual correlation is calculated; this is, it is not the conditional correlation, and the bandwidth does not correspond to an exact number of observations. Figure 1 shows the estimated local Gaussian correlation plotted out on a grid chosen by the user, and I have pointed this out in the first line of the paragraph that introduces this figure.

Although one can discuss the "effective number of observations" in the classical nonparametric sense, this has not been done in the literature on the LGC, and hence not a consept in the *lg*-package.

> *5. Is it possible to test for significant differences in the correlation coefficients, for instance, from the lower left, middle and upper right in Fig. 1?*

My response: That may certainly be possible. To the best of my knowledge, however, this is not a test that has been introduced in the literature. It is therefore not implemented in the **lg**-package.

> *6. Page 1: I suggest to change the statement 'clever and creative techniques' by a more scientific formulation.*

My response: This statement has been taken out. Thank you.