# The locally Gaussian partial correlation

Håkon Otneim[*]and Dag Tjøstheim[†]

**Abstract**

It is well known in econometrics and other fields that the dependence structure for jointly Gaussian variables can be fully captured using correlations, and that the *conditional* dependence structure in the same way can be described using partial correlations. The partial correlation does not, however, characterize conditional dependence in many non-Gaussian populations. This paper introduces the local Gaussian partial correlation (LGPC), a new measure of conditional dependence. It is a local version of the partial correlation coefficient that characterizes conditional dependence in a large class of populations. It has some useful and novel properties besides: The LGPC reduces to the ordinary partial correlation for jointly normal variables, and it distinguishes between positive and negative conditional dependence. Furthermore, the LGPC can be used to study departures from conditional independence in specific parts of the distribution. We provide several examples of this, both simulated and real, and derive estimation theory under a local likelihood estimation framework. Finally, we indicate how the LGPC can be used to construct a powerful test for conditional independence, which, for example, can be used to detect nonlinear Granger causality in time series.

**Keywords:** *Conditional dependence, conditional independence test, Granger causality*

[*]Corresponding author. Department of Business and Management Science, NHH Norwegian School of Economics. Helleveien 30, 5045 Bergen, Norway. `hakon.otneim@nhh.no`

[†]University of Bergen

# 1  INTRODUCTION

Estimation of conditional dependence and testing for conditional independence are important topics in classical as well as modern statistics. In the last two decades, for instance, there have been intense developments in using conditional dependence for probabilistic networks, conditional multivariate time series analysis as well as copula analysis. For jointly Gaussian variables, conditional dependence is measured by the partial correlation coefficient. Given three jointly Gaussian stochastic variables $X_1$, $X_2$ and $X_3$, $X_1$ and $X_2$ are conditionally independent given $X_3$ if and only if the partial correlation between $X_1$ and $X_2$ given $X_3$ is equal to zero. There is a rich literature on applications of the partial correlation coefficient to Gaussian networks, path analysis and causality. The Gaussian assumption is strict, however, and it is easy to find non-Gaussian examples where the partial correlation function completely fails in describing conditional dependence, as will be seen later in this article.

The purpose of this paper is to introduce a new concept for measuring conditional dependence. This concept retains all the properties of the ordinary partial correlation in the Gaussian case, but seeks to avoid the weaknesses of this measure in the non-Gaussian case. This is done by fitting a *family* of Gaussian distributions to a given continuous multivariate distribution and exploiting the simple conditioning rules of the Gaussian distribution *locally*. This approach produces a new measure of conditional dependence, *the local Gaussian partial correlation (LGPC)* which, being directly related to the ordinary partial correlation, is easy to interpret, and reduces to the very same ordinary partial correlation in the Gaussian case. Moreover, it distinguishes between positive and negative conditional dependence whereas competing non-linear measures report only the *strength* of the conditional dependence on some non-negative scale.

The local view gives much more flexibility. It allows the conditional dependence to be stronger or weaker in certain regions of a multivariate distribution than in others. This is of particular interest in finance, where description of tail behavior is important. The local aspect also makes it possible to focus tests of conditional independence to selected parts of

the distribution, which may potentially increase the power of the test.

The local approach has been shown to be advantageous in other areas of statistics, such as the measurement of nonlinear dependence, density estimation and spectral analysis, see for instance Tjøstheim and Hufthammer (2013), Otneim and Tjøstheim (2017) and Jordanger and Tjøstheim (2020). The present paper, then, represents the first attempt to model conditional dependence locally. There are several other approaches to modelling conditional dependence and testing for conditional independence in the non-Gaussian case in the literature. For the most part, they revolve around test statistics in conditional independence tests. Such statistics are usually based on measures of distance between some property of the sample and the corresponding property of the population under the null hypothesis of conditional independence. As a consequence, the conditional dependence measures are not always easy to interpret, and can, as mentioned above, not distinguish between negative and positive conditional dependence. We present a comprehensive set of references to this literature in Section 5.

Definitions and properties are presented in Section 2, whereas estimation is covered in Section 3. Two examples illustrating the strengths of our approach are given in Section 4. Section 5 gives a brief account of the potential in testing.

An integral part of our article is the online supplement. It deals with a number of issues complementing the treatment in the main paper. Among them are more examples, extensions of the main concept to more general situations, precise formulations of some theoretical results with proofs, as well as discussions of local asymptotic power and the Pitman criterion for testing and validity of the use of the bootstrap.

# 2 THE LOCAL GAUSSIAN PARTIAL CORRELATION

Let $\boldsymbol{X} = (X_1, \ldots, X_p)$ be a random vector. Denote by $(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3)$ a partition of $\boldsymbol{X}$ into vectors of dimension $d_1$, $d_2$ and $d_3$ respectively, such that $\boldsymbol{X}^{(1)} = (\boldsymbol{X}_1, \boldsymbol{X}_2) = (X_1, \ldots, X_{d_1+d_2})$ consists of the first $d_1 + d_2$ components in $\boldsymbol{X}$ and $\boldsymbol{X}^{(2)} = \boldsymbol{X}_3 = (X_{d_1+d_2+1}, \ldots, X_{d_1+d_2+d_3})$ contains the remaining $d_3$ variables, where $p = d_1 + d_2 + d_3$. We assume that the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{X}$ exist, and partition them correspondingly, writing

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \\ \boldsymbol{\mu}_3 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{13} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \boldsymbol{\Sigma}_{23} \\ \boldsymbol{\Sigma}_{31} & \boldsymbol{\Sigma}_{32} & \boldsymbol{\Sigma}_{33} \end{pmatrix}, \tag{1}$$

where $\boldsymbol{\Sigma}_{ij}$ is the covariance matrices of $(\boldsymbol{X}_i, \boldsymbol{X}_j)$, $i, j = 1, 2, 3$. There are two main concepts of correlation when $\boldsymbol{X}^{(2)} = \boldsymbol{X}_3$ is given, the *partial* and the *conditional* correlation, and they coincide in several joint distributions, among them the Gaussian. See e.g. Baba, Shibata, and Sibuya (2004) for details. We will use the partial correlation as a starting point when defining the LGPC. The partial variance-covariance matrix of $\boldsymbol{X}^{(1)} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$ given $\boldsymbol{X}^{(2)} = \boldsymbol{X}_3$ is

$$\boldsymbol{\Sigma}_{12|3} = \boldsymbol{\Sigma}^{11} - \boldsymbol{\Sigma}^{12} \left(\boldsymbol{\Sigma}^{22}\right)^{-1} \boldsymbol{\Sigma}^{21}, \tag{2}$$

where

$$\boldsymbol{\Sigma}^{11} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad \boldsymbol{\Sigma}^{12} = \begin{pmatrix} \boldsymbol{\Sigma}_{13} \\ \boldsymbol{\Sigma}_{23} \end{pmatrix}, \quad \boldsymbol{\Sigma}^{21} = \begin{pmatrix} \boldsymbol{\Sigma}_{31} & \boldsymbol{\Sigma}_{32} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}^{22} = \boldsymbol{\Sigma}_{33},$$

and $\boldsymbol{\Sigma}_{12|3}$ is the covariance matrix in the conditional (Gaussian) distribution of $\boldsymbol{X}^{(1)}$ given $\boldsymbol{X}_3$ if $\boldsymbol{X}$ is jointly normal. The partial correlation matrix between $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ given $\boldsymbol{X}_3$ is naturally defined as

$$\boldsymbol{R}_{12|3} = \boldsymbol{D}^{-1/2}\boldsymbol{\Sigma}_{12|3}\boldsymbol{D}^{-1/2}, \tag{3}$$

where $\boldsymbol{D} = \operatorname{diag}(\boldsymbol{\Sigma}_{12|3})$. We identify in the same way the partial correlation matrix (3) with the correlation matrix in the conditional (Gaussian) distribution of $\boldsymbol{X}^{(1)}$ given $\boldsymbol{X}^{(2)}$ if $\boldsymbol{X}$ is jointly normal. Equations (2) and (3) will serve as the starting point for our definition of the *local* partial correlation.

## 2.1 Definition

In all of the following we assume that the components of $\boldsymbol{X}$ are continuous with joint density function $f_{\boldsymbol{X}}$. Given a point $\boldsymbol{x}$, we approximate $f_{\boldsymbol{X}}$ in a neighborhood of $\boldsymbol{x}$ by a multivariate Gaussian density

$$\psi(\boldsymbol{x}, \boldsymbol{v}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}(\boldsymbol{x})|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{v} - \boldsymbol{\mu}(\boldsymbol{x}))^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x})(\boldsymbol{v} - \boldsymbol{\mu}(\boldsymbol{x}))\right\}, \tag{4}$$

where $\boldsymbol{x} = (x_1 \ldots, x_p)$, $\boldsymbol{\mu}(\boldsymbol{x}) = \{\mu_j(\boldsymbol{x})\}$ and $\boldsymbol{\Sigma}(\boldsymbol{x}) = \{\sigma_{jk}(\boldsymbol{x})\}$ for $j, k = 1, \ldots, p$. Moving to another point $\boldsymbol{y}$, there is another (generally different) Gaussian approximation $\psi(\boldsymbol{y}, \boldsymbol{v})$. In this way we approximate $f_{\boldsymbol{X}}$ by a family of multivariate Gaussian densities defined by a set of smooth parameter functions $\{\boldsymbol{\mu}(\boldsymbol{x}), \boldsymbol{\Sigma}(\boldsymbol{x})\}$, and if $f_{\boldsymbol{X}}$ is itself a Gaussian density, then the parameter functions collapse to constants corresponding to the true parameter values, and $\psi(\boldsymbol{x}) \equiv f_{\boldsymbol{X}}(\boldsymbol{x})$. Hjort and Jones (1996) provide the general framework for estimating such parameter functions non-parametrically from a given data set using a local likelihood procedure, and the basic idea in the following treatment is to replace the components in the partial covariance matrix (2) by their locally estimated counterparts in order to obtain a local measure of conditional dependence.

Before we get that far, however, we will introduce a useful transformation technique that improves and simplifies the estimation of the LGPC. Otneim and Tjøstheim (2017, 2018) show that the estimation of the local parameter functions $\{\boldsymbol{\mu}(\boldsymbol{x}), \boldsymbol{\Sigma}(\boldsymbol{x})\}$ becomes

easier by transforming each $X_j$ to a standard normal variable $Z_j = \Phi^{-1}(U_j)$, where $U_j$ is a uniform variable $U_j = F_j(X_j)$ with $F_j$ being the cumulative distribution function of $X_j$. Define the random vector $\boldsymbol{Z}$ by this transformation of $\boldsymbol{X} = (\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}) = (\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3) = (X_1, X_2, \ldots, X_p)$ to marginal standard normality:

$$\boldsymbol{Z} = \left( \Phi^{-1}\left(F_1(X_1)\right), \Phi^{-1}\left(F_2(X_2)\right), \ldots, \Phi^{-1}\left(F_p(X_p)\right) \right). \tag{5}$$

The transformation enables us to simplify the local Gaussian approximation (4) by writing the density $f_{\boldsymbol{Z}}$ of $\boldsymbol{Z}$ at the point $\boldsymbol{v} = \boldsymbol{z}$ as

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \psi(\boldsymbol{z}, R(\boldsymbol{z})) = \frac{1}{|2\pi \boldsymbol{R}(\boldsymbol{z})|^{1/2}} \exp\left\{ -\frac{1}{2} \boldsymbol{z}^T \boldsymbol{R}^{-1}(\boldsymbol{z}) \boldsymbol{z} \right\}, \tag{6}$$

where, as in Otneim and Tjøstheim (2017, 2018) in a further simplified approximation, we have fixed local means and standard deviations $\mu_j(\boldsymbol{z}) = 0$ and $\sigma_j^2(\boldsymbol{z}) = 1$, $j = 1, \ldots, p$, and where $\boldsymbol{R}(\boldsymbol{z}) = \{\rho_{jk}(\boldsymbol{z})\}$ is the *local correlation matrix*.

In practice, we do not know $F_j$, but we can instead use the empirical distribution function

$$\widehat{F}_j(x) = \frac{1}{n-1} \sum_{i=1}^{n} 1\left(X_{ji} \leq x\right),$$

where $1(\cdot)$ is the indicator function, $n$ is the number of observations, and $X_{ji}$ is the $i$th observation of $X_j$. This results in pseudo standard normal variables $\widehat{Z}_j = \Phi^{-1}(\widehat{F}_j(X_j))$.

In the following, we will not always distinguish between $Z_j$ and $\widehat{Z}_j$. In fact, we conclude our asymptotic analysis in Section 3.2 by showing that under certain technical conditions, the error made by estimating $\boldsymbol{R}(\boldsymbol{z})$ using the empirically transformed variables $Z_{j,n}$, instead of $Z_j$, is smaller in the limit than the estimation error made when estimating the local correlations themselves.

We will in this paper refer to $\boldsymbol{X}$ and its probability density function $f_{\boldsymbol{X}}$ as being on the

$x$-scale, and to $\boldsymbol{Z}$ and its probability density function $f_{\boldsymbol{Z}}$ as being on the $z$-scale. For further discussion of the simplified $z$-approximation we refer to Otneim and Tjøstheim (2017).

Denote by $(\boldsymbol{Z}^{(1)}, \boldsymbol{Z}^{(2)}) = (\boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{Z}_3)$ the partitioning of $\boldsymbol{Z}$ corresponding to the partitioning $(\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}) = (\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3)$ of $\boldsymbol{X}$. A natural definition of the *local* partial covariance matrix of $\boldsymbol{Z}^{(1)}|\boldsymbol{Z}^{(2)}$ is the local version of eq. (2).

$$\boldsymbol{\Sigma}_{12|3}(\boldsymbol{z}) = \boldsymbol{R}^{11}(\boldsymbol{z}^{(1)}) - \boldsymbol{R}^{12}(\boldsymbol{z}) \left( \boldsymbol{R}^{22}(\boldsymbol{z}^{(2)}) \right)^{-1} \boldsymbol{R}^{21}(\boldsymbol{z}). \tag{7}$$

If $d_1 = d_2 = 1$ then $\boldsymbol{\Sigma}_{12|3}(\boldsymbol{z})$ is a $2 \times 2$ matrix, and we define the local Gaussian partial correlation $\alpha(\boldsymbol{z})$ between the two variables in $\boldsymbol{Z}^{(1)} = (Z_1, Z_2)$ given $\boldsymbol{Z}^{(2)} = Z_3$ in accordance with the ordinary (global) partial correlation provided by eq. (3):

$$\alpha(\boldsymbol{z}) = R_{12|3}(\boldsymbol{z}) = \frac{\left\{ \boldsymbol{\Sigma}_{12|3}(\boldsymbol{z}) \right\}_{12}}{\left\{ \boldsymbol{\Sigma}_{12|3}(\boldsymbol{z}) \right\}_{11}^{1/2} \left\{ \boldsymbol{\Sigma}_{12|3}(\boldsymbol{z}) \right\}_{22}^{1/2}}, \tag{8}$$

which, if $\boldsymbol{Z}^{(2)} = Z_3$ is scalar, reduces to

$$\alpha(\boldsymbol{z}) = \rho_{12|3}(z_1, z_2|z_3) = \frac{\rho_{12}(z_1, z_2) - \rho_{13}(z_1, z_3)\rho_{23}(z_2, z_3)}{\sqrt{1 - \rho_{13}^2(z_1, z_3)}\sqrt{1 - \rho_{23}^2(z_2, z_3)}}. \tag{9}$$

It is of course possible to introduce an LGPC $\alpha(\boldsymbol{x})$ directly on the $x$-scale, but that representation is in many ways harder to handle both computationally and asymptotically. For a multivariate Gaussian distribution, we have that $\alpha_{\boldsymbol{X}}(\boldsymbol{x}) = \alpha_{\boldsymbol{Z}}(\boldsymbol{z}) = \alpha$. In the remainder of this paper, we write mainly in terms of the $z$-representation using the LGPC $\alpha(\boldsymbol{z}) = \alpha_{\boldsymbol{Z}}(\boldsymbol{z})$, but when we write the local partial correlation between $X_1$ and $X_2$ given $\boldsymbol{X}_3 = \boldsymbol{x}_3$ at the point $(x_1, x_2, \boldsymbol{x}_3)$, this is simply $\alpha(\boldsymbol{z})$ inserted $z_j = \Phi^{-1}(F_j(x_j))$, $j = 1, \ldots, p$.

In the more general case where $d_1 > 1$ and/or $d_2 > 1$ then $\boldsymbol{\Sigma}_{12|3}(\boldsymbol{z})$ is a $(d_1 + d_2) \times (d_1 + d_2)$ matrix that describes the non-linear conditional dependence among the variables in $\boldsymbol{Z}^{(1)} = (\boldsymbol{Z}_1, \boldsymbol{Z}_2)$ given $\boldsymbol{Z}^{(2)} = \boldsymbol{Z}_3$ (or alternatively $\boldsymbol{X}^{(1)} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$ given $\boldsymbol{X}^{(2)} = \boldsymbol{X}_3$), which in particular can be used to analyze the conditional dependence between two *sets* of random

variables having dimensions $d_1$ and $d_2$ respectively, given a third set of variables having dimension $d_3$. This case poses no conceptual challenges to our approach, but it requires a rather large investment in new notation, and does not lead to simple formulas like eq. (9). We will therefore, for the most part, focus on the local Gaussian partial correlation between two stochastic scalar variables in the remainder of this paper. We provide the complete description of the general multivariate case in Sections A.1 and A.2 of the online supplement.

## 2.2 Properties

The local Gaussian partial correlation is closely related to the partial correlation between jointly normally distributed variables. We see this also from the following properties that we will establish next:

1. The LGPC $\alpha(\boldsymbol{z})$ satisfies $-1 \leq \alpha(\boldsymbol{z}) \leq 1$.
2. If $\boldsymbol{X}$ is jointly normally distributed, then the LGPC coincides with the ordinary (global) partial correlation.
3. For stochastic vectors having joint density function on $z$-scale of the form (6), the LGPC $\alpha(\boldsymbol{z})$ is identically equal to zero if and only if $X_1$ and $X_2$ are conditionally independent given $\boldsymbol{X}_3$. Note that $\alpha(\boldsymbol{z}) \equiv 0$ if and only if $\alpha(\boldsymbol{x}) \equiv 0$.
4. The LGPC $\alpha(\boldsymbol{z})$ is invariant with regard to a set of monotone transformations $\boldsymbol{Y} = \boldsymbol{h}(\boldsymbol{X}) = (h_1(X_1), \ldots, h_p(X_p))$.

Property 1 is trivially true if $\boldsymbol{R}(\boldsymbol{z})$ is a valid correlation matrix. By removing the $\boldsymbol{z}$-dependence in the local correlations, it follows immediately from eq. (6) and from the results referred to in Section 2.1 that property 2 holds.

To see that conditional independence between $X_1$ and $X_2$ given $\boldsymbol{X}_3$ is equivalent to $\alpha(\boldsymbol{z}) \equiv 0$, or equivalently $\alpha(\boldsymbol{x}) \equiv 0$, we need to follow a few simple steps that would also work in the global Gaussian case. Working on the standard normal $z$-scale and assuming that $(z_1, z_2, \boldsymbol{z}_3)$ has density function $f_{\boldsymbol{Z}}(\boldsymbol{z})$ as given by (6), it follows from Otneim and Tjøstheim

8

(2018) that we can calculate conditional distributions in the same way as in the global Gaussian case. For example, the conditional density of $Z_1|\boldsymbol{Z}_3$ at the point $z_1$ is given by

$$f_{z_1|z_3}(z_1|\boldsymbol{z}_3) = \frac{1}{|2\pi\sigma_{1|3}(z_1)|^{1/2}} \times \exp\left\{-\frac{1}{2}\left(\frac{z_1 - \mu_{1|3}(z_1)}{\sigma_{1|3}(z_1)}\right)^2\right\},$$

where

$$\mu_{1|3}(z_1) = \boldsymbol{R}_{12}(z_1, \boldsymbol{z}_3)\boldsymbol{R}_{22}(\boldsymbol{z}_3)^{-1}\boldsymbol{z}_3 \quad \text{and} \quad \sigma_{1|3}(z_1) = \boldsymbol{\Sigma}_{Z_1|\boldsymbol{Z}_3}, \tag{10}$$

and where the latter expression is defined analogously to (7). The conditional density $f_{Z_2|\boldsymbol{Z}_3}(z_2|\boldsymbol{z}_3)$ is defined in the same way. If $Z_1$ and $Z_2$ are conditionally independent given $\boldsymbol{Z}_3$, then

$$\begin{aligned}
f_{Z_1,Z_2|\boldsymbol{Z}_3}(z_1, z_2|\boldsymbol{z}_3) &= f_{Z_1|\boldsymbol{Z}_3}(z_1|\boldsymbol{z}_3)f_{Z_2|\boldsymbol{Z}_3}(z_2|\boldsymbol{z}_3) \\
&= \frac{1}{2\pi\sqrt{1 - \sigma_{1|3}(z_1)}\sqrt{1 - \sigma_{2|3}(z_2)}} \\
&\quad \times \exp\left\{-\frac{1}{2}\left(\frac{z_1 - \mu_{1|3}(z_1)}{\sigma_{1|3}(z_1)}\right)^2\right\}\exp\left\{-\frac{1}{2}\left(\frac{z_2 - \mu_{2|3}(z_2)}{\sigma_{2|3}(z_2)}\right)^2\right\},
\end{aligned}$$

and we at once identify the conditional density of $(Z_1, Z_2)|\boldsymbol{Z}_3$ as another Gaussian distribution, but without any cross-term involving $z_1z_2$, which then implies that the off-diagonal element in the partial covariance matrix (7) is identically equal to zero. We see immediately from our definition (8) that the LGPC is also identically equal to zero.

The converse statement follows straightforwardly by considering the conditional density of $(Z_1, Z_2|\boldsymbol{Z}_3)$.

Let the transformation $\boldsymbol{Y} = \boldsymbol{h}(\boldsymbol{X})$ define the $y$-scale in the same way as the transformation (5) defines the $z$-scale. We define the LGPC in terms of the marginally standard normal $\boldsymbol{Z}$-variables, which for the $\boldsymbol{Y}$-variables can be calculated as

$$\boldsymbol{Z} = \left( \Phi^{-1}\left( F_{Y_1}(Y_1) \right), \ldots, \Phi^{-1}\left( F_{Y_p}(Y_p) \right) \right),$$

but if $h_j(\cdot)$ is monotone for all $j = 1, \ldots, p$,

$$F_{Y_j}(y_j) = \mathbb{P}(h_j(X_j) \leq y_j) = \mathbb{P}(X_j \leq h_j^{-1}(y_j)) = \mathbb{P}(X_j \leq x_j) = F_{X_j}(x_j),$$

where $x_j$ is the point on the $x$-scale corresponding to the point $y_j$ on the $y$-scale. The $\boldsymbol{Z}$-variables are the same for the stochastic variables $\boldsymbol{X}$ and $\boldsymbol{Y} = \boldsymbol{h}(\boldsymbol{X})$. Their LGPC-function $\alpha(\boldsymbol{z})$ must therefore be the same as well according to our definition in the preceding section. Entirely analogous arguments for the properties 1.-4. in the general multivariate case can be constructed from the supplementary material, Sections A.1 and A.2.

# 3   ESTIMATING CONDITIONAL DEPENDENCE USING THE LGPC

## 3.1   Estimation by local likelihood

From the defining equations (7) and (8) it is seen that the basic building blocks for the *local Gaussian partial correlation* are the *local Gaussian correlation functions*, that populate the local correlation matrix $\boldsymbol{R}(\boldsymbol{z})$ in the density function $f_{\boldsymbol{Z}}$ in (6). Consider the random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)$ having joint probability density function (pdf) $f(\boldsymbol{x})$, and its transformed counterpart $\boldsymbol{Z} = (\Phi^{-1}(F_1(X_1)), \ldots, \Phi^{-1}(F_p(X_p)))$, on the marginally standard normal $z$-scale. The relation between the pdf of $\boldsymbol{X}$ and the pdf of $\boldsymbol{Z}$ is given by

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = f_{\boldsymbol{Z}}\left( \Phi^{-1}\left( F_1(x_1) \right), \ldots, \Phi^{-1}\left( F_p(x_p) \right) \right) \prod_{j=1}^{p} \frac{f_j(x_j)}{\phi\left( \Phi^{-1}\left( F_j(x_j) \right) \right)}, \qquad (11)$$

where $f_j(\cdot)$, $j = 1, \ldots, p$ are the marginal density functions of $\boldsymbol{X}$, and $\phi(\cdot)$ is the standard normal density function. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a random sample identically distributed as $\boldsymbol{X}$

and construct the pseudo standard normal observations $\widehat{\boldsymbol{Z}}_1, \ldots, \widehat{\boldsymbol{Z}}_n$ as

$$\widehat{\boldsymbol{Z}}_t = \left( \Phi^{-1}\left(\widehat{F}_1(X_{1,t})\right), \ldots, \Phi^{-1}\left(\widehat{F}_p(X_{p,t})\right)\right), \tag{12}$$

where $\boldsymbol{X}_t = (X_{1,t}, X_{2,t}, \ldots, X_{p,t})$, and $\widehat{F}_j(\cdot)$ is an estimate of the marginal distribution function of $X_j$, for example the empirical distribution function. Next, we must produce an estimate $\widehat{\boldsymbol{R}}(\boldsymbol{z})$ of the local correlation matrix, and we will consider two variations of this task in subsections 3.1.1 and 3.1.2 with a more general treatment in Section A.3 of the supplement. It is then, following (7), natural to estimate the local Gaussian partial covariance matrix as

$$\widehat{\boldsymbol{\Sigma}}_{12|3}(\boldsymbol{z}) = \widehat{\boldsymbol{R}}^{11}(\boldsymbol{z}) - \widehat{\boldsymbol{R}}^{12}(\boldsymbol{z})\left(\widehat{\boldsymbol{R}}^{22}(\boldsymbol{z})\right)^{-1}\widehat{\boldsymbol{R}}^{21}(\boldsymbol{z}), \tag{13}$$

and if we assume that $X_1$ and $X_2$ constitute the two first elements in $\boldsymbol{X}$, then, following (8), the local Gaussian partial correlation between $X_1$ and $X_2$ given $\boldsymbol{X}^{(2)} = \boldsymbol{X}_3$ on the $z$-scale is estimated by

$$\widehat{\alpha}(\boldsymbol{z}) = \frac{\left\{\widehat{\boldsymbol{\Sigma}}_{12|3}(\boldsymbol{z})\right\}_{12}}{\left\{\widehat{\boldsymbol{\Sigma}}_{12|3}(\boldsymbol{z})\right\}_{11}^{1/2}\left\{\widehat{\boldsymbol{\Sigma}}_{12|3}(\boldsymbol{z})\right\}_{22}^{1/2}}. \tag{14}$$

A corresponding value of $\widehat{\alpha}(\boldsymbol{x})$ at the point $\boldsymbol{x} = \widehat{F}^{-1}(\Phi(z))$ is obtained by inserting $\boldsymbol{z} = \Phi^{-1}(\widehat{F}(\boldsymbol{x}))$ in (14).

### 3.1.1 Estimation of $\boldsymbol{R}(\boldsymbol{z})$ when $p = 3$ and $\boldsymbol{X}^{(2)}$ is a scalar

If $\dim(\boldsymbol{X}) = 3$, then $\dim(\boldsymbol{X}^{(2)}) = 1$, and $\boldsymbol{R}(\boldsymbol{z})$ is a $3 \times 3$ symmetric matrix of local correlation functions having arguments $\boldsymbol{z} = (z_1, z_2, z_3)$:

$$\boldsymbol{R}(\boldsymbol{z}) = \begin{pmatrix} 1 & \rho_{12}(\boldsymbol{z}) & \rho_{13}(\boldsymbol{z}) \\ \rho_{12}(\boldsymbol{z}) & 1 & \rho_{23}(\boldsymbol{z}) \\ \rho_{13}(\boldsymbol{z}) & \rho_{23}(\boldsymbol{z}) & 1 \end{pmatrix}, \tag{15}$$

11

containing three parameter functions that must be estimated from data. Note that here we deviate from the pairwise scheme found in Otneim and Tjøstheim (2017, 2018) as well as most of this article where $\rho_{ij}(\boldsymbol{z}) = \rho_{ij}(z_i, z_j)$. In equation (15), $\boldsymbol{z} = (z_1, z_2, z_3)$. Based on the sample (12), $\boldsymbol{R}(\boldsymbol{z})$ is estimated by fitting the parametric family $\psi(z, \boldsymbol{R})$, defined in (6), locally to the density $f_{\boldsymbol{Z}}(\boldsymbol{z})$ by maximizing the *local likelihood function* (cf. Hjort and Jones (1996)),

$$\widehat{\boldsymbol{R}}(\boldsymbol{z}) = \arg\max_{\boldsymbol{R}} n^{-1} \sum_{t=1}^{n} K_{\boldsymbol{b}}(\widehat{\boldsymbol{Z}}_t - \boldsymbol{z}) \log \psi(\widehat{\boldsymbol{Z}}_t, \boldsymbol{R}) - \int K_{\boldsymbol{b}}(\boldsymbol{y} - \boldsymbol{z}) \psi(\boldsymbol{y}, \boldsymbol{R}) \, d\boldsymbol{y} \qquad (16)$$

in each point $\boldsymbol{z}$, where $K_{\boldsymbol{b}}(\boldsymbol{x}) = |\boldsymbol{b}|^{-1} K(\boldsymbol{b}^{-1}\boldsymbol{x})$, $K$ is a non-negative and radially symmetric kernel function that satisfies $\int K(\boldsymbol{x}) \, d\boldsymbol{x} = 1$, and $\boldsymbol{b}$ is a diagonal $3 \times 3$ matrix of bandwidths that serve as smoothing parameters. The estimate $\widehat{f}_{\boldsymbol{Z}}(\boldsymbol{z}) = \psi(\boldsymbol{z}, \widehat{\boldsymbol{R}}(\boldsymbol{z}))$ aims at minimizing a locally weighted Kullback-Leibler distance to $f_{\boldsymbol{Z}}(\boldsymbol{z})$, and we refer to Hjort and Jones (1996) and Tjøstheim and Hufthammer (2013) for much more details about this construction, and to Section 3.2 for the relevant asymptotic estimation theory.

Finally, we obtain the estimated LGPC $\widehat{\alpha}(\boldsymbol{z})$ by plugging the estimated local correlations $\widehat{\boldsymbol{R}}(\boldsymbol{z})$ into equations (13) and (14) and a corresponding value of $\widehat{\alpha}(\boldsymbol{x})$ at the point $\boldsymbol{x} = \widehat{F}^{-1}(\Phi(\boldsymbol{z}))$ by inserting $\boldsymbol{z} = \Phi^{-1}(\widehat{F}(\boldsymbol{x}))$.

### 3.1.2 Estimation of $\boldsymbol{R}(\boldsymbol{z})$ when $\boldsymbol{X}^{(2)}$ is a vector

The complexity of the estimation problem (16) increases sharply with the number of variables involved. The basic idea of Otneim and Tjøstheim (2017) to avoid the curse of dimensionality is to model local correlations in $\boldsymbol{R}(\boldsymbol{z})$ as functions of their corresponding *pair* of variables only:

$$\boldsymbol{R}(\boldsymbol{z}) = \{\rho_{jk}(z_j, z_k)\}, \qquad j, k = 1, 2, \ldots, p, \qquad (17)$$

12

which reduces the estimation of the $p$-variate correlation functions $\rho_{jk}(\boldsymbol{z})$ to a series of bivariate estimation problems. We suggest to use this approach when modelling local dependence if the number of variables in $\boldsymbol{X}$ is larger than three. This allows nonlinear dependence between variables to be approximated by the pairwise structure (17), while still being computationally tractable. An analogue to the pairwise approximation is the additive approximation in nonparametric regression.

We transform the observations to the $z$-scale as in (12), but now estimate the individual components in $\boldsymbol{R}(\boldsymbol{z})$ one by one, taking only into consideration the *pair* of variables in question:

$$\widehat{\rho}_{jk}(z_j, z_k) = \arg\max_{\rho_{jk}} n^{-1} \sum_{t=1}^{n} K_{\boldsymbol{b}}(\widehat{\boldsymbol{Z}}_t - \boldsymbol{z}) \log \psi(\widehat{\boldsymbol{Z}}_t, \rho_{jk}) - \int K_{\boldsymbol{b}}(\boldsymbol{y} - \boldsymbol{z})\psi(\boldsymbol{y}, \rho_{jk}) \, \mathrm{d}\boldsymbol{y}, \quad (18)$$

where all running variables and samples are bivariate subsets corresponding to the indices $(j, k)$, $\psi(\cdot, \rho)$ is the bivariate version of (6), and $\boldsymbol{b}$ now is a $2 \times 2$ diagonal matrix of bandwidths. After estimating all local correlations in this way, we proceed to calculate the LGPC using equations (13) and (14) as above.

## 3.2   Asymptotic theory

Equations (7), (8) and their empirical counterparts (13) and (14) demonstrate clearly that the LGPC is nothing more than a deterministic function of the local correlation matrix, in the same way as the ordinary partial correlation is a function of the ordinary correlation matrix. The asymptotic behavior of $\widehat{\alpha}(\boldsymbol{z})$ can thus be derived directly from the asymptotic behavior of $\widehat{\boldsymbol{R}}(\boldsymbol{z})$, which, for the most part, has been established in earlier works, see e.g. Tjøstheim and Hufthammer (2013) and Otneim and Tjøstheim (2017, 2018). Some key results follow below, and we refer to the supplementary material, Sections B–D for proofs of the asymptotic distributional results in this subsection, and to Section A.3 for a discussion of more general situations and for a result on consistency.

We start by assuming that the marginal distribution functions $F_1, \ldots, F_p$ are known, and then state two results on the joint asymptotic normality of the estimated local correlation matrix $\widehat{\boldsymbol{R}}(\boldsymbol{z})$ under a mixing condition. We will then apply the delta method in order to derive the limiting distribution of the estimated LGPC, and finally show that we may replace $F_1, \ldots, F_p$ with their empirical counterparts $\widehat{F}_1, \ldots, \widehat{F}_p$ without changing these asymptotic results.

Consider first the full trivariate fit $\widehat{\boldsymbol{R}}(\boldsymbol{z})$ described in Section 3.1.1 with $\boldsymbol{z} = (z_1, z_2, z_3)$. We estimate the local correlation matrix $\boldsymbol{R}(\boldsymbol{z})$ by maximizing the local likelihood function in (16), which we denote by $L_n(\boldsymbol{R}(\boldsymbol{z}), \boldsymbol{z})$, with the only exception that we, for now, assume the true transformation between the $x$- and the $z$-scale to be known. For a fixed matrix of bandwidths $\boldsymbol{b}$, denote by $\boldsymbol{R_b}(z)$ the local correlation that satisfies, as $n \to \infty$,

$$\nabla L_n(\boldsymbol{R_b}, \boldsymbol{z}) \to \int K_{\boldsymbol{b}}(\boldsymbol{y} - \boldsymbol{z}) \boldsymbol{u}(\boldsymbol{y}, \boldsymbol{R_b}(\boldsymbol{z}))\{f_{\boldsymbol{Z}}(\boldsymbol{y}) - \psi(\boldsymbol{y}, \boldsymbol{R_b})\} \, \mathrm{d}\boldsymbol{y} = 0, \tag{19}$$

where $f_{\boldsymbol{Z}}(\boldsymbol{z}) = f_{Z_1, Z_2, Z_3}(z_1, z_2, z_3)$ is the density function of $\boldsymbol{Z}$ (at the point $\boldsymbol{z}$), and $\boldsymbol{u}(\cdot, \rho_b)$ is the column vector of local score functions $\nabla \log \psi(\cdot, \rho_b)$, and where the gradient is taken with respect to the parameters $\rho_{12}(\boldsymbol{z})$, $\rho_{13}(\boldsymbol{z})$ and $\rho_{23}(\boldsymbol{z})$. Denote by $\boldsymbol{\rho} = \boldsymbol{\rho}(\boldsymbol{z}) = \{\rho_{12}(\boldsymbol{z}), \rho_{13}(\boldsymbol{z}), \rho_{23}(\boldsymbol{z})\}$ the vector of local correlations. The joint limiting distribution of $\widehat{\boldsymbol{\rho}}$ (for convenience only stated on the $z$-scale below) is given by the following result as $n \to \infty$ and $\boldsymbol{b} \to \boldsymbol{0}$:

**Theorem 3.1.** *Let $\{\boldsymbol{Z}_t\}_{t=1}^n$ be observations on the marginally standard normally distributed random vector $\boldsymbol{Z}$ having joint density $f_{\boldsymbol{Z}}(\boldsymbol{z})$. Assume that the following conditions hold:*

1. *For any sequence of bandwidth matrices $\boldsymbol{b}_n$ tending to zero element-wise, there exists for the trivariate marginally standard Gaussian vector $\boldsymbol{Z}$ a unique set of local correlations $\boldsymbol{\rho_b}(\boldsymbol{z})$ that satisfies (19), and there exists a $\boldsymbol{\rho}_0(\boldsymbol{z})$ such that $\boldsymbol{\rho_b}(\boldsymbol{z}) \to \boldsymbol{\rho}_0(\boldsymbol{z})$.*

2. *The process generating $\{\boldsymbol{Z}_t\}_{t=1}^n$ is $\alpha$-mixing with mixing coefficients $\alpha(m)$ satisfying $\sum_{m \geq 1} m^\lambda \alpha(m)^{1-2/\delta} < \infty$ for some $\lambda > 1 - 2/\delta$ and $\delta > 2$.*

3. $n \to \infty$, and each bandwidth $b = b_n$ tends to zero such that $nb^{\frac{\lambda+2-2/\delta}{\lambda+2/\delta}} = O(n^{\epsilon_0})$ for some constant $\epsilon_0 > 0$.

4. In a given point $\boldsymbol{z}$, the parameter space $\Theta$ for each local correlation $\rho(\boldsymbol{z})$ is a compact subset of $(-1, 1)$.

5. The non-negative kernel function $K(\cdot)$ satisfies $\sup_{\boldsymbol{z}} K(\boldsymbol{z}) < \infty$, $\int K(\boldsymbol{y})\, d\boldsymbol{y} < \infty$, $\partial/\partial z_i K(\boldsymbol{z}) < \infty$, and $\lim_{z_i \to \infty} |z_i K(\boldsymbol{z})| = 0$ for $i = 1, 2, 3$.

Then, by writing $b_{n1} = b_{n2} = b_{n3} = b \to 0$ assuming they all converge to zero at the same rate,

$$\sqrt{nb^3} \boldsymbol{M}_{\boldsymbol{b}}^{-1/2} \boldsymbol{J}_{\boldsymbol{b}} \left(\widehat{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_0\right) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}), \tag{20}$$

where $\boldsymbol{I}$ is the $3 \times 3$ identity matrix,

$$\boldsymbol{J}_{\boldsymbol{b}} = \int K_b(\boldsymbol{y} - \boldsymbol{z}) \boldsymbol{u}(\boldsymbol{y}, \boldsymbol{\rho}_b(\boldsymbol{z})) \boldsymbol{u}^T(\boldsymbol{y}, \boldsymbol{\rho}_b(\boldsymbol{z})) \psi(\boldsymbol{y}, \boldsymbol{\rho}_b(\boldsymbol{z}))\, d\boldsymbol{y}$$
$$- \int K_b(\boldsymbol{y} - \boldsymbol{z}) \nabla \boldsymbol{u}(\boldsymbol{y}, \boldsymbol{\rho}_b(\boldsymbol{z})) \Big\{ f_{\boldsymbol{Z}}(\boldsymbol{y}) - \psi(\boldsymbol{y}, \boldsymbol{\rho}_b(\boldsymbol{z})) \Big\} d\boldsymbol{y},$$

and

$$\boldsymbol{M}_{\boldsymbol{b}} = b^3 \int K_b^2(\boldsymbol{y} - \boldsymbol{z}) \boldsymbol{u}(\boldsymbol{y}, \boldsymbol{\rho}_b(\boldsymbol{z})) \boldsymbol{u}^T(\boldsymbol{y}, \boldsymbol{\rho}_b(\boldsymbol{z})) f_{\boldsymbol{Z}}(\boldsymbol{y})\, d\boldsymbol{y}$$
$$- b^3 \int K_b(\boldsymbol{y} - \boldsymbol{z}) \boldsymbol{u}(\boldsymbol{y}, \boldsymbol{\rho}_b(\boldsymbol{z})) f_{\boldsymbol{Z}}(\boldsymbol{y})\, d\boldsymbol{y} \int K_b(\boldsymbol{y} - \boldsymbol{z}) \boldsymbol{u}^T(\boldsymbol{y}, \boldsymbol{\rho}_b(\boldsymbol{z})) f_{\boldsymbol{Z}}(\boldsymbol{y})\, d\boldsymbol{y}.$$

It is shown in Section B of the Supplement that the matrices $\boldsymbol{J}_{\boldsymbol{b}}$ and $\boldsymbol{M}_{\boldsymbol{b}}$ are both of order $b^2$. Theorem 3.1 and its assumptions above are variations of the assumptions and the results presented by Hjort and Jones (1996), and in particular Theorem 3 by Tjøstheim and Hufthammer (2013). The analysis of the asymptotic behavior of $\boldsymbol{J}_{\boldsymbol{b}}$ and $\boldsymbol{M}_{\boldsymbol{b}}$ in the latter reference reveals that the local parameter estimates converge more slowly than, for example, the trivariate kernel density estimator (that converges as $1/(nb^3)^{1/2}$), and it is a straightforward exercise to modify their proof to suit our particular situation, which demonstrates that the three local correlations converge jointly in distribution at the rate

15

$1/(nb^5)^{1/2}$, see Section B in the supplement for details.

Furthermore, Tjøstheim and Hufthammer (2013) show that an analytic expression of the leading term in the covariance matrix $\boldsymbol{J}_{\boldsymbol{b}}^{-1}\boldsymbol{M}_b(\boldsymbol{J}_{\boldsymbol{b}}^{-1})^T$ is not practically available. In applications we may circumvent this problem by using the bootstrap. The validity of the bootstrap can be proved using a direct adaptations of the techniques of Lacal and Tjøstheim (2018). It is spelled out in some detail for the pairwise case in Theorems E.1, E.2 and corollaries E.3 and E.4 in Section E of the online supplement. The same arguments can be used in the present 3-parameter case (in fact, Lacal and Tjøstheim (2018) treat the more general five-parameter case).

The pairwise estimation structure that we describe in Section 3.1.2 simplifies the asymptotic analysis of $\widehat{\boldsymbol{R}}(\boldsymbol{z})$ considerably. The result is essentially available in the literature as Theorem 2 in Otneim and Tjøstheim (2018), who state under similar assumptions as in Theorem 3.1 that

$$\sqrt{nb_n^2}\left(\widehat{\boldsymbol{\rho}}(\boldsymbol{z}) - \boldsymbol{\rho}_0(\boldsymbol{z})\right) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Lambda}), \tag{21}$$

where the asymptotic covariance matrix $\boldsymbol{\Lambda}$ is defined as $\boldsymbol{\Sigma}$ in Theorem 2 of Otneim and Tjøstheim (2018). There are two important differences compared to Theorem 3.1, however. First of all, we have that the convergence rate in the simplified pairwise case is equal to the usual nonparametric rate of $1/(nb^2)^{1/2}$. This is because we estimate the local correlations one by one based on pairs of variables. The second difference is that the leading term of the asymptotic covariance matrix in Theorem 2 by Otneim and Tjøstheim (2018) has a simple analytic expression that can be used to construct, for instance, confidence intervals. We refer to Section A.3 in the online supplement for further details regarding the limiting distribution of the LGPC as well conditions for consistency of the estimates.

We can now calculate the limiting distribution of the LGPC by means of the delta method. Denote by $g : \mathbb{R}^{p(p-1)/2} \to \mathbb{R}$ the translation (8) from the vector of local correlations between the components in $\boldsymbol{Z}$ (or, equivalently, $\boldsymbol{X}$) to the local partial correlation $\alpha(\boldsymbol{z})$ between $Z_1$ and $Z_2$ given $Z_3, \ldots, Z_p$. From the expression of our estimate (14) we have that $\widehat{\alpha}(\boldsymbol{z}) = g(\widehat{\boldsymbol{\rho}}(\boldsymbol{z}))$,

so we use the delta method to see that

$$\sqrt{nb_n^m}\left(\widehat{\alpha}(\boldsymbol{z}) - \alpha(\boldsymbol{z})\right) \xrightarrow{d} \mathcal{N}\left(0, \nabla g(\boldsymbol{\rho})^T \boldsymbol{\Lambda} \nabla g(\boldsymbol{\rho})\right), \tag{22}$$

where $m$ is equal to 2 or 5, and $\boldsymbol{\Lambda}$ is either equal to the leading term of $\boldsymbol{J_b}^{-1}\boldsymbol{M_b}(\boldsymbol{J_b}^{-1})^T$, or $\boldsymbol{\Lambda}$ in (21), depending on whether we use the full trivariate locally Gaussian fit described in Section 3.1.1 or the sequentially pairwise simplification described in Section 3.1.2. In any case, it is only a matter of basic differentiation to work out an expression for $\nabla g(\cdot)$, and this task has been deferred to Section C of the online supplement in the case where $d_1 = d_2 = 1$.

Finally, we present the limiting distribution of the LGPC between $Z_1$ and $Z_2$ given $Z_3, \ldots, Z_p$ when the components in $\boldsymbol{Z} = (Z_1, \ldots, Z_p)$ are replaced by marginally standard normal *pseudo-observations* as defined in equation (12). The following result ensures that Theorems 3.1 and Theorem 2 by Otneim and Tjøstheim (2018), and thus eq. (22), still holds.

**Theorem 3.2.** *Under the conditions in Theorems 3.1 and Theorem 2 by Otneim and Tjøstheim (2018), assume further that the kernel function $K(\cdot)$ has bounded support. Replacing the marginally standard normal vector $\boldsymbol{Z}$ with the approximately marginally standard normal vector $\widehat{\boldsymbol{Z}}$, does not change the conclusions in these theorems.*

*Proof.* See Section D of the online supplement. $\square$

# 4   EXAMPLES

We will in this section provide some demonstrations on how the estimated LGPC can be used to reveal nonlinear departures from conditional independence and to measure that conditional dependence. All plots are presented on the $x$-scale.

First, we provide some reference pictures from the simplest situation imaginable. Let $X_1$, $X_2$ and $X_3$ be independent standard (and thus also jointly) normal variables. Since $X_1$ is conditionally independent from $X_2$ given $X_3$ (written $X_1 \perp X_2|X_3$) in this particular case,

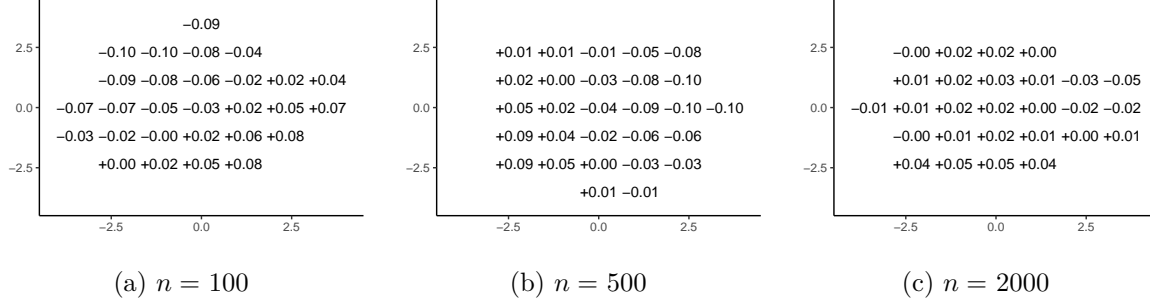we also know from property 3 in Section 2.2 that the LGPC between $X_1$ and $X_2$ is zero everywhere.



Figure 1: Estimated local Gaussian partial correlation between $X_1$ and $X_2$ given $X_3 = 0$, where $(X_1, X_2, X_3)$ is a jointly standard normally distributed vector. The fully trivariate model of Section 3.1.1 is used.

In Figure 1 we see the estimated LGPC between $X_1$ and $X_2$ given $X_3 = 0$ for three samples, having sample size 100, 500 and 2000 respectively, mapped out on a grid. In this and all other examples we select bandwidths based on a simple plug-in formula that follows naturally from classical asymptotic arguments, $b = cn^{-1/9}$ for the full trivariate fit, and $b = cn^{-1/6}$ for the bivariate simplification, where the constant $c$ controls the amount of smoothing and must be chosen appropriately based on the task at hand. Otneim (2016) argues that $c = 1.75$ is a good choice for multivariate density estimation, and we will in the next section see that we obtain good power in our test for conditional independence if $c$ is somewhat smaller than that. For the visual display of conditional dependence maps, however, we tend to prefer a fair amount of smoothing, and in Figure 1 we have used $c = 4$. Otneim and Tjøstheim (2017) also suggest a algorithm for bandwidth selection in the density estimation context, maximizing the leave-one-out cross validated log-likelihood of the sample;

$$\boldsymbol{b}_{CV} = \arg \max_{\boldsymbol{b}} \sum_{t=1}^{n} \widehat{f}_{\boldsymbol{b}}^{(-t)}(\boldsymbol{Z}_t),$$

where $\widehat{f}_{\boldsymbol{b}}^{(-t)}(\boldsymbol{Z}_t)$ is the estimated local Gaussian density estimate calculated without, and evaluated in, the sample point $\boldsymbol{Z}_t$, using the bandwidth matrix $\boldsymbol{b}$. A theory of efficient

and accurate bandwidth selection for local likelihood estimation remains a topic for future research.

As expected, $\widehat{\alpha}(\boldsymbol{x}) = \widehat{\alpha}(\boldsymbol{z})$ is close to zero in all three plots in Figure 1 and a significance test carried out as in Section 5 does not reject conditional independence at any reasonable significance level.

In the next simulated example we will demonstrate an important feature of the LGPC: It is able to *distinguish between positive and negative conditional relationships*, which, to our knowledge, has until now not been possible beyond the linear and jointly Gaussian setting using the ordinary partial correlation. Our approach also allows exploration of conditionally different dependence patterns across different levels of the conditioning variable.

Generate $X_3 = \rho \sim U(-1, 1)$, and then generate $(X_1, X_2)$ from the bivariate Gaussian distribution having standard normal marginals, and correlation coefficient equal to $\rho$. We observe $\boldsymbol{X} = (X_1, X_2, \rho)$ and seek to visualize the dependence between $X_1$ and $X_2$ conditional on $\rho$.

We see the results in Figure 2. In panel (a), we display $n = 1000$ simulated pairs $(X_1, X_2)$ from this model. In panel (b)-(d) we see the estimated LGPC plotted over suitable grids, and where the conditioning variable $X_3 = \rho$ has been fixed at the respective values $-0.9$, $0$ and $0.9$, and we see clearly how the dependence between $X_1$ and $X_2$ dramatically changes in an intuitively reasonable way. By a conditioning argument one can see that $X_1$ and $X_2$ are uncorrelated and that the ordinary partial correlation between $X_1$ and $X_2$ given $X_3$ is equal to zero. Furthermore, we note that the pairwise simplification defined in Section 3.1.2 would also not be able to measure the *conditional* relationship between $X_1$ and $X_2$ given $X_3$ in this case, as we see clearly from eq. (9), because $X_1$ and $X_2$ are both marginally independent from $X_3$. This means that the two pairwise correlations $\rho_{13}(z_1, z_3)$ and $\rho_{23}(z_2, z_3)$ are equal to zero.
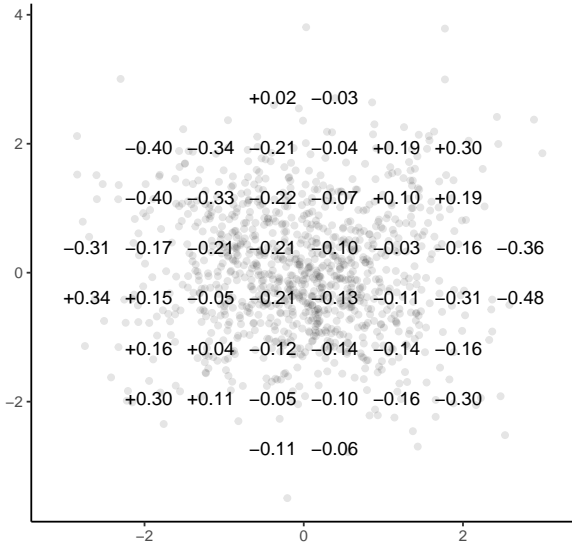
We conclude this section by using the LGPC to analyze real data in an example regarding Granger causality between financial time series. Granger causality has been a central concept
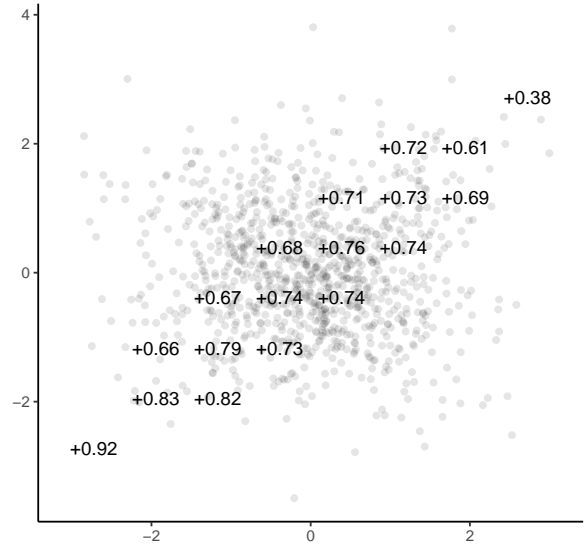
19
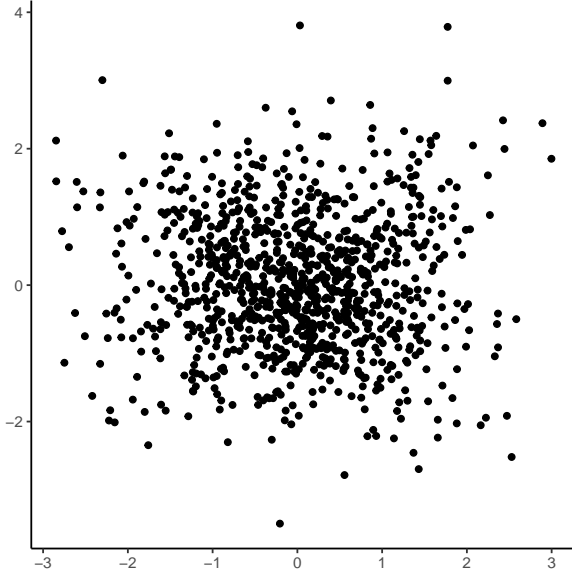
(a) Observed pairs $(X_1, X_2)$ from the conditionally Gaussian model.
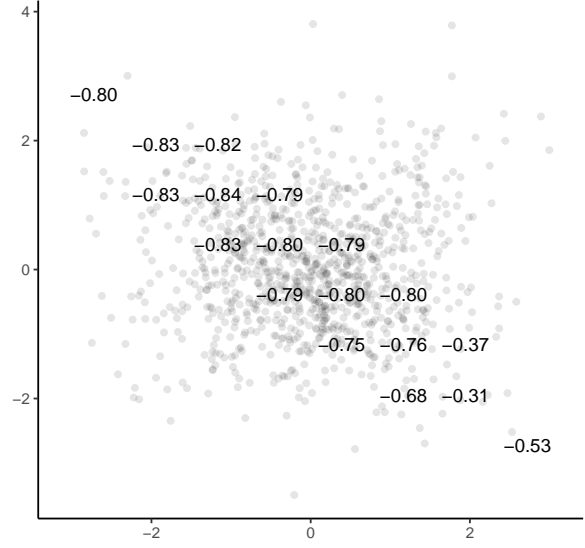
(b) The estimated LGPC between $X_1$ and $X_2$ given $\rho = -0.9$.

(c) The estimated LGPC between $X_1$ and $X_2$ given $\rho = 0$

(d) The estimated LGPC between $X_1$ and $X_2$ given $\rho = 0.9$

Figure 2: The conditionally Gaussian model

in economics and econometrics ever since its inception by Granger (1969). In layman's terms, the time series $\{X_t\}$ Granger causes $\{Y_t\}$ if past values of $\{X_t\}$ are helpful when predicting future values of $\{Y_t\}$. Formally, $\{X_t\}$ Granger causes $\{Y_t\}$ if (Granger 1980)

$$Y_t \not\perp \mathcal{I}^*(t-1) \mid \mathcal{I}^*_{-X}(t-1), \tag{23}$$

where $\perp$ denotes independence, $\not\perp$ denotes dependence, and where $\mathcal{I}^*(t-1)$ is all information available at time $t-1$ and $\mathcal{I}^*_{-X}(t-1)$ is the same information, with the exception of the values of $\{X_t\}$ up to, but not including, time $t$. Of course, the hypothesis (23) can not be tested in practice in its full generality. By taking only effects up to the first lag into account, we may formulate a sufficient (but not necessary) condition for (23): $Y_t \not\perp X_{t-1} \mid Y_{t-1}$, the converse of which constitutes a testable null hypothesis of Granger non-causality:

$$\mathrm{H}_0 : Y_t \perp X_{t-1} \mid Y_{t-1}. \tag{24}$$

There are many ways to carry out this test. The simplest, and original one, is based on the further restriction of linear relationships between $\{X_t\}$ and $\{Y_t\}$ and is thus a test for *conditional uncorrelatedness* rather than conditional independence. Nonparametric tests that have power against many nonlinear types of conditional dependence have also been developed, and we refer to Section 5 for further references to this literature and a new test for conditional independence based on the LGPC.

Francis, Mougoue, and Panchenko (2010) investigate how stock returns of large firms may Granger cause the stock returns of small firms and vice versa from a nonlinear standpoint. They point to earlier evidence that such nonlinear lead-lag relationships have been found and investigate this further using the test for Granger non-causality by Baek and Brock (1992), and using a long series of US data they find evidence that there is a bi-directional causal link between the stock prices of the firms in the most valuable quintile (20 percent) and the firms in the least valuable quintile. We will repeat this exercise for stocks traded on the Oslo Stock

Exchange in Norway, but this time from a point of view using the estimated LGPC.

Denote by $R_{1,t}$ the log-returns on a value-weighted portfolio consisting of the most valuable quintile of firms listed on the Oslo Stock Exchange as of 27 April 2020 (the big firms), and denote by $R_{2,t}$ the corresponding log-returns for the least valuable companies (the small firms). We will investigate the possibility of Granger causality from the small to the big firms, the hypothesis being that smaller firms are more market sensitive and quickly react to changes in the economy, leading the larger firms that more slowly adapt to small variations. The null hypothesis of non-causality of the first order (24) in this particular case means that

$$R_{1,t} \perp R_{2,t-1} \mid R_{1,t-1}, \tag{25}$$

which we may test formally using the approaches described in the following section or the references therein. Let us rather concentrate on *describing* this conditional dependence relationship by estimating the local Gaussian partial correlation between $R_{1,t}$ and $R_{2,t-1}$ given a few different values of $R_{2,t-1}$, which gives considerable more information than a single $p$-value of a test. We collect daily observations on Oslo Stock Exchange[1] from 1 January 2018 until 31 December 2019. In Figure 3 we see the estimated LGPC along the diagonal $R_{1,t} = R_{2,t-1}$ for three different values $R_{1,t-1} = -0.02, 0$ and 0.02, along with their block-bootstrapped 95% confidence intervals. The constant $\alpha(\boldsymbol{z}) = 0$ is indicated as a dotted line.

One could perhaps make the argument that $R_{1,t}$ and $R_{2,t-1}$ seem to be somewhat positively dependent in the right half of the graph for *small* values of the conditioning variable $R_{1,t-1}$. Roughly translated to this particular situation this means that if the large-firm portfolio gives a negative return on day $t - 1$, then its value on day $t$ depend on the performance small-firm portfolio on day $t - 1$ in a nonlinear and non-Gaussian way that is characterized by positive dependence in the first quadrant. The evidence displayed in Figure 3 is quite weak, however,

---

[1]Thanks to *Børsprosjektet (The Stock Exchange Project)* at the Norwegian School of Economics for providing the data.
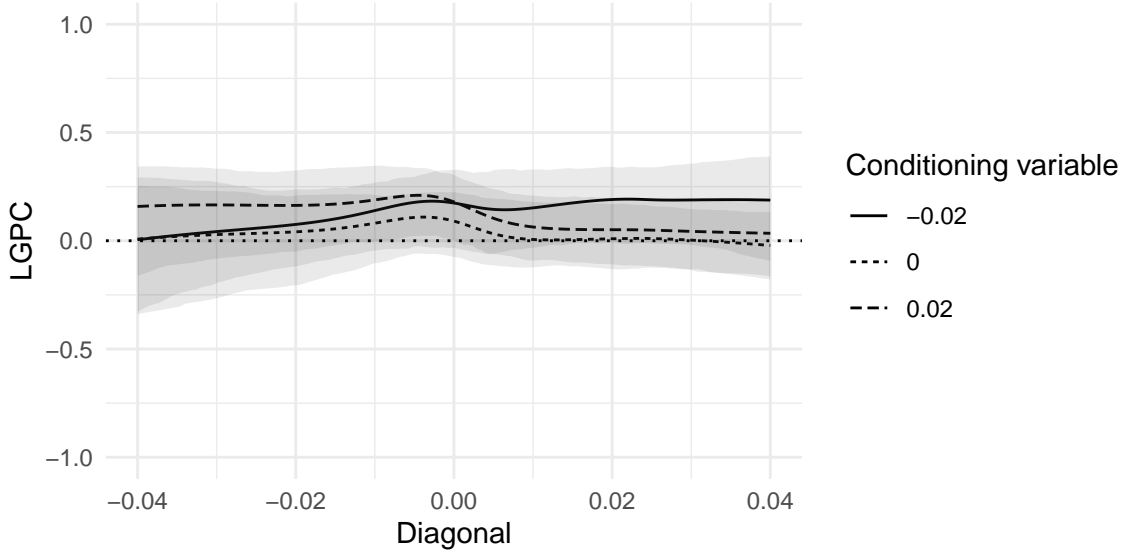
Figure 3: The estimated LGPC between $R_{1,t}$ and $R_{2,t-1}$ given $R_{1,t-1}$ along the diagonal of the sample space, for three values of the conditioning variable

because the confidence intervals include zero in most of the sample space.

Francis, Mougoue, and Panchenko (2010) go on to analyze if there is Granger causality from small to big firms in the US data if another lag is taken into account, which in general means finding evidence against the hypothesis

$$R_{1,t} \perp (R_{2,t-1}, R_{2,t-2}) \,|\, (R_{1,t-1}, R_{1,t-2}), \tag{26}$$

departures from which can be quantified and tested for in a linear fashion using multiple regressions, or tested for directly by means of a nonlinear test as we discuss in the following Section. Let us here, on the other hand, use the LGPC to quantify the conditional dependence relationship (26) nonlinearly for the Norwegian returns data in order to obtain more detailed information. This means estimating a $3 \times 3$ partial correlation matrix $\hat{\boldsymbol{\alpha}}(\boldsymbol{z})$, which we again do along the diagonal $R_{1,t} = R_{2,t-1} = R_{2,t-2}$ in this particular case.

Including another lag does not change the estimated conditional relationship between $R_{1,t}$ and $R_{2,t-1}$ much compared to the first order estimate displayed in Figure 3. There does, however, seem to be a clearer picture when considering the LGPC between $R_{1,t}$ and
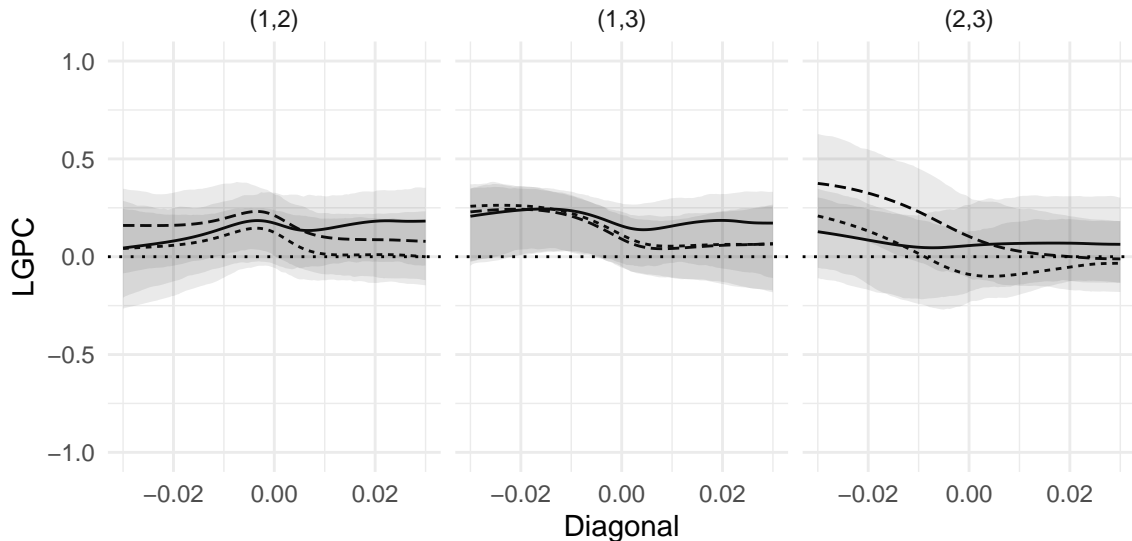
Figure 4: The estimated local Gaussian partial correlations for the three variable pairs in $(R_{1,t}, R_{2,t-1}, R_{2,t-2})$ given $(R_{1,t-1}, R_{1,t-2})$ estimated along the diagonal in the sample space. The first two plots represent the pairs $(R_{1,t}, R_{2,t-1})$ and $(R_{1,t}, R_{2,t-2})$ which are of particular interest when analyzing the conditional dependence relationship (26). The legend in Figure 3 is still valid, except that the solid line now represents the LGPC conditional on *both* $R_{1,t-1}$ and $R_{1,t-2}$ being equal to $-0.02$.

$R_{2,t-2}$, which seems to be positive in the third quadrant for all values of the conditioning variables. In this particular problem this means that conditioning on the log return of the small-firm portfolio on days $t-1$ and $t-2$, the large-firm log-return on day $t$ depends on the small-firm log-return on day $t-2$ in a nonlinear way, and more specifically; this conditional relationship can be described using the LGPC as stronger dependence in the lower left tail of the relevant conditional distribution. We present another example of Granger testing in addition to another simulated experiment in Section J of the online supplement.

# 5    TESTING FOR CONDITIONAL INDEPENDENCE

We should point out here that the purpose of the LGPC is not primarily to compete with the nonparametric conditional independence *testing* literature, but rather to provide a new

tool for the measurement of nonlinear conditional dependence. The following section could however be seen as an indication of the potential of using the LGPC to test for conditional independence. It remains to be investigated whether further improvements can be realized by using a 5-parameter specification and/or a focused version of the test; see e.g. Tables 3–5 of Lacal and Tjøstheim (2017) for the substantial improvements in a similar focused independence test.

We restrict ourselves to the case where $X_1$ and $X_2$ are scalars, since this is the situation considered in the papers we compare with, but we reiterate (end of Section 2.1) that there are no conceptual difficulties in extending to the vector case as is seen in the online supplement. The theory underlying LGPC testing is essentially based on Lacal and Tjøstheim (2018), and is outlined in Sections F–H of the supplement.

## 5.1   The recent fauna of nonparametric tests

Property 3 in Section 2.2 states that the LGPC characterizes conditional dependence within a quite large class of distributions: Two stochastic vectors $X_1$ and $X_2$ are independent given $\boldsymbol{X}_3$ if and only if the locally Gaussian partial correlation between them is identically equal to zero everywhere on the sample space of $(X_1, X_2, \boldsymbol{X}_3)$. The road is therefore short to a test for conditional independence that may have power against a great deal of nonlinear alternatives compared to a test based on the ordinary partial correlation coefficient.

The last decade or so has seen the publication of many new tests for conditional independence. Su and White have published a series of such tests: Su and White (2007) is based on detecting differences between estimated characteristic functions (which is also the method used by Wang and Hong (2018)). Su and White (2008) is based on estimating the Hellinger distance between conditional density estimates, Su and White (2012) use local polynomial quantile regression to test for conditional independence, and Su and White (2014) present a test based on empirical likelihood. Huang (2010) introduces the maximal nonlinear conditional correlation which is used in a test for conditional independence, in turn extended

to dependent data by Cheng and Huang (2012). Song (2009) constructs a test via Rosenblatt transformations, while Bergsma (2010) and Bouezmarni, Rombouts, and Taamouti (2012) present new tests for conditional independence based on copula constructions. Bouezmarni and Taamouti (2014) test for conditional independence by measuring the $L_2$ distance between estimated conditional distribution functions, and Patra, Sen, and Székely (2016) use empirical transformations to translate conditional independence to joint independence, tests for which exist in abundance. Wang et al. (2015) introduce the *conditional distance correlation* which they use to construct a test for conditional independence. Linton and Gozalo (1997, 2014) formulate conditional independence in terms of probability statements which then forms the basis of a test. Also, Delgado and Manteiga (2001) develop a test for conditional independence in a nonparametric regression framework. Finally, we mention testing by way of characterizing conditional independence via reproducing kernel Hilbert spaces (RKHS), see, for example, Zhang et al. (2012) and references therein.

A test based on the LGPC is quite different from the methods quoted above. It is semi-parametric and does not rely on traditional kernel estimates of density-, distribution-, or characteristic functions. Furthermore, due to our transformation of the data (5) to marginal standard normality, we do not necessarily have to specify weight functions in our test functional to lessen the impact of outliers - unless of course we wish to test for conditional independence in a specific portion of the sample space.

On the other hand, we have to pay a price when imposing structure on the dependence. Certain types of conditional dependence will remain invisible to our test statistic, but simulation experiments show that our test performs on par with, and sometimes better than existing, fully nonparametric tests.

## 5.2   A test for conditional independence based on the LGPC

We construct a test statistic for testing $H_0 : X_1 \perp X_2 \mid \boldsymbol{X}_3$, or, equivalently, a test statistic in terms of the marginally Gaussian pseudo observations,

$$\mathrm{H}_0 : Z_1 \perp Z_2 \mid \boldsymbol{Z}_3, \tag{27}$$

by aggregating our local measure of dependence over the sample space of $\boldsymbol{X}$ (or $\boldsymbol{Z}$). A natural test statistic on the $z$-scale is

$$T_{n,b} = \int_S h\left(\widehat{\alpha}_b(\boldsymbol{z})\right) \, \mathrm{d}F_n(\boldsymbol{z}), \tag{28}$$

where $h(\cdot)$ is a real-valued function that is typically even and non-negative for most standard applications, and $S \subseteq \mathbb{R}^p$ is an integration area that can be altered in order to focus the test to specific portions of the sample space. Standard laws of large numbers ensure that, under regularity conditions, see Theorem F.1 of the supplement, $T_{n,b}$ converges in probability towards its population value

$$T = \int_S h\left(\alpha(\boldsymbol{z})\right) \, \mathrm{d}F(\boldsymbol{z}).$$

Departures from conditional independence lead to large values of $T_{n,b}$ that, if larger than a critical value, leads to the rejection of (27). Asymptotically, one might expect that approximate $p$-values for our test can be extracted from the limiting distribution of the test statistic $T_{n,b}$, which we derive in the online supplement Section F. The limiting asymptotic distribution of $T_{n,b} - T$ is given in Theorem F.2 for the 1-parameter general pairwise case, whereas Theorem F.3 treats the 3-coordinate (3-parameter) case with $d_1 = d_2 = d_3 = 1$. It is interesting to note from the discussion immediately following the proofs of these theorems that the convergence rate in the first case is $n^{-1/2}$, and in the second case is $(nb^2)^{-1/2}$. This means that the tests are asymptotically able to pick up alternatives approaching the null hypothesis of conditional independence at these rates. This is measured by the Pitman criterion. A discussion of the Pitman criterion for other tests of conditional independence is given on pages 11–12 immediately following Theorem 2 of Wang and Hong (2018). The rate $n^{-1/2}b^{-d_3/4}$ of Wang and Hong compares favorable with other tests. The rate $n^{-1/2}$ obtained by us must be interpreted with caution, however, because it is based on a pairwise model

that is more restrictive. We go into more detail regarding local asymptotic power and the Pitman criterion in Section G of the online supplement.

Several authors, for example Teräsvirta, Tjøstheim, and Granger (2010), have noted, however, that asymptotic analysis of nonparametric (or semiparametric) test statistics on the form (28) tends to be too crude for finite sample applications. We have therefore approximated the distribution of $T_{n,b}$ using the bootstrap. The validity of the bootstrap in such contexts has been examined by Lacal and Tjøstheim (2018) and we adapt their arguments to our situation in the online supplement, Section H.

In accordance with our treatment so far, let $\{Z_{1,t}, Z_{2,t}, \boldsymbol{Z}_{3,t}\}$, $t = 1, \ldots, n$ be observations on the $p$-variate stochastic vector $\boldsymbol{Z}$, $p \geq 3$, with $Z_1$ and $Z_2$ being scalar and $\boldsymbol{Z}_3$ being $(p-2)$-variate. In order to calculate the statistic $T_{n,b}$ for testing the null hypothesis $Z_1 \perp Z_2 \mid \boldsymbol{Z}_3$, we must estimate the joint local correlation matrix $\boldsymbol{\Sigma}(\boldsymbol{z})$ of $\boldsymbol{Z} = (Z_1, Z_2, \boldsymbol{Z}_3)$. We exploit this in the following algorithm designed to produce approximate resampled versions of $T_{n,b}$ under the null hypothesis of conditional independence. This procedure is a variation of the so-called local bootstrap which has been used by several authors when testing for conditional independence. A brief survey of this procedure is given in Section I of the online supplement.

1. Use the local correlation estimates to estimate the conditional densities $f_{Z_1|\boldsymbol{Z}_3}(\cdot)$ and $f_{Z_2|\boldsymbol{Z}_3}(\cdot)$ by means of the method by Otneim and Tjøstheim (2018). In practice, in order to reduce the computational load, we evaluate $\widehat{f}_{Z_1|\boldsymbol{Z}_3}$ and $\widehat{f}_{Z_2|\boldsymbol{Z}_3}$ on a fine grid on their support, over which a continuous representation of the estimates are produced using cubic splines.

2. Using the accept-reject algorithm, generate $B$ samples, each of size $n$, from $\widehat{f}_{Z_1|\boldsymbol{Z}_3}$ and $\widehat{f}_{Z_2|\boldsymbol{Z}_3}$, leading to $B$ replicates

$$Z_m^* = \left\{\{Z_{1t}^*, Z_{2t}^*, \boldsymbol{Z}_{3t}^*\}_i\right\}_m, \qquad t = 1, \ldots, n, \ m = 1, \ldots, B,$$

of $\boldsymbol{Z}$ under the null hypothesis.

| | |
|---|---|
| 1. | $(X_{1,t}, X_{2,t}, X_{3,t}) = (\epsilon_{1,t}, \epsilon_{2,t}, \epsilon_{3,t})$. |
| 2. | $X_{1,t} = 0.5X_{1,t-1} + \epsilon_{1,t},\ X_{2,t} = 0.5X_{2,t-1} + \epsilon_{2,t},\ X_{3,t} = X_{1,t-1}$. |
| 3. | $X_{1,t} = \epsilon_{1,t}\sqrt{0.01 + 0.5X_{1,t-1}^2},\ X_{2,t} = 0.5X_{2,t-1} + \epsilon_{2,t},\ X_{3,t} = X_{1,t-1}$. |
| 4. | $X_{1,t} = \epsilon_{1,t}\sqrt{h_{1,t}},\ X_{2,t} = \epsilon_{2,t}\sqrt{h_{2,t}},\ X_{3,t} = X_{1,t-1}$, |
| | $h_{1,t} = 0.01 + 0.9h_{1,t-1} + 0.05X_{1,t-1}^2,\ h_{2,t} = 0.01 + 0.9h_{2,t-1} + 0.05X_{2,t-1}^2$. |
| 5. | $X_{1,t} = 0.5X_{1,t-1} + 0.5X_{2,t} + \epsilon_{1,t},\ X_{2,t} = 0.5X_{2,t-1} + \epsilon_{2,t},\ X_{3,t} = X_{1,t-1}$. |
| 6. | $X_{1,t} = 0.5X_{1,t-1} + 0.5X_{2,t}^2 + \epsilon_{1,t},\ X_{2,t} = 0.5X_{2,t-1} + \epsilon_{2,t},\ X_{3,t} = X_{1,t-1}$. |
| 7. | $X_{1,t} = 0.5X_{1,t-1}0.5X_{2,t} + \epsilon_{1,t},\ X_{2,t} = 0.5X_{2,t-1} + \epsilon_{2,t},\ X_{3,t} = X_{1,t-1}$. |
| 8. | $X_{1,t} = 0.5X_{1,t-1} + 0.5X_{2,t}\epsilon_{1,t},\ X_{2,t} = 0.5X_{2,t-1} + \epsilon_{2,t},\ X_{3,t} = X_{1,t-1}$. |
| 9. | $X_{1,t} = \epsilon_{1,t}\sqrt{0.01 + 0.5X_{1,t-1}^2 + 0.25X_{2,t}},\ X_{2,t} = X_{2,t-1} + \epsilon_{2,t},\ X_{3,t} = X_{1,t-1}$. |
| 10. | $X_{1,t} = \epsilon_{1,t}\sqrt{h_{1,t}},\ X_{2,t} = \epsilon_{2,t}\sqrt{h_{2,t}},\ X_{3,t} = X_{1,t-1},\ h_{1,t} = 0.01 + 0.1h_{1,t-1} + 0.4X_{1,t-1}^2 +$ |
| | $0.5X_{2,t}^2,\ h_{2,t} = 0.01 + 0.9h_{2,t-1} + 0.5X_t^2$. |

Table 1: Data generating processes in simulation experiment

3. Calculate $\{T_{n,b}\}_m^*$, $m = 1, \ldots, B$ for the replicated data sets and obtain the approximate $p$-value for the conditional independence test.

## 5.3 Comparing with other tests

Su and White (2008) formulate a simulation experiment for evaluating their nonparametric test for conditional independence by generating data from 10 different data generating processes (DGPs), and then check the power and level for their test. Many of the later works that were discussed in Section 5.1 contain very similar experiments, and some of them present simulation results for tests published prior to Su and White (2008). This allows us to present comprehensive comparisons between the new test presented above, and many alternatives.

Let $(\epsilon_{1,t}, \epsilon_{2,t}, \epsilon_{3,t})$ be IID observations from a $\mathcal{N}(0, I_3)$-distribution, where $I_3$ is the $3 \times 3$ identity matrix. We will test $\mathrm{H}_0 : X_{1,t} \perp X_{2,t} \mid X_{3,t}$ in 10 cases listed in Table 1 taken from Su and White (2008), that cover various types of linear and nonlinear time series dependence.

The null hypotheses of conditional independence between $X_{1,t}$ and $X_{2,t}$ given $X_{3,t}$ is true for DGPs 1–4, and these will be used to check the level of the test, while for DGPs 5–10 we measure the power. By evaluating our test using the sample sizes $n = 100$ and $n = 200$ at the 5% level, we can harvest a great number of corresponding results from the literature, and

| | Level | | | | Power | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| ↓ Test \| DGP → | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| CHF | 0.034 | 0.058 | – | – | 0.780 | 0.792 | 0.520 | 0.780 | 0.728 | 0.580 |
| CM | 0.054 | 0.058 | 0.060 | 0.048 | 0.920 | 0.548 | 0.504 | 0.412 | 0.384 | 0.188 |
| HEL, $c=2$ | 0.072 | 0.036 | 0.072 | 0.048 | 0.952 | <u>0.944</u> | 0.576 | 0.940 | <u>0.988</u> | <u>0.912</u> |
| KS | 0.042 | 0.056 | 0.056 | 0.040 | 0.780 | 0.404 | 0.380 | 0.288 | 0.292 | 0.156 |
| LGPC, $c=1.0$ | 0.054 | 0.048 | 0.046 | 0.046 | 0.910 | 0.722 | 0.559 | <u>0.990</u> | 0.968 | 0.866 |
| LGPC, $c=1.4$ | 0.047 | 0.043 | 0.046 | 0.047 | 0.971 | 0.855 | 0.727 | 0.969 | 0.916 | 0.765 |
| LIN | 0.044 | 0.061 | 0.050 | 0.060 | <u>0.999</u> | 0.337 | 0.213 | 0.126 | 0.163 | 0.153 |
| MCC, $c=2$ | 0.041 | 0.050 | 0.053 | 0.062 | 0.852 | 0.793 | 0.218 | 0.860 | 0.631 | 0.348 |
| SCM | 0.076 | 0.060 | 0.084 | 0.064 | 0.924 | 0.464 | 0.352 | 0.500 | 0.224 | 0.196 |
| SEL | 0.054 | 0.038 | – | – | 0.840 | 0.856 | <u>0.760</u> | 0.904 | 0.716 | 0.556 |
| SKS | 0.064 | 0.056 | 0.088 | 0.068 | 0.728 | 0.236 | 0.288 | 0.340 | 0.120 | 0.112 |

Table 2: Level and power, $n = 100$

they are presented below.

The first set of simulation results are collected from Su and White (2008): In Table 2 LIN is a standard linear Granger causality test, SCM and SKS are tests by Linton and Gozalo (1997) that use statistics of the two versions of the nonparametric test developed by Delgado and Manteiga (2001). HEL refers to the Hellinger distance test that Su and White (2008) present. Next, we move to Su and White (2014), who provide simulations for their test for conditional independence based on the empirical likelihood (SEL), as well as the test by Su and White (2007) that is based on properties of the conditional characteristic function (CHF). Cheng and Huang (2012) provide simulations for their tests based on the maximal conditional correlation (MCC). In the online supplement to this article we provide a richer version of Table 2 below, namely Table 2*, that includes more results for the HEL- and MCC-tests at different levels of smoothing.

Finally, we include results from simulations using our new test based on the LGPC, using the trivariate specification defined in Section 3.1.1, for two different levels of smoothing, and include them in Table 2. We highlight the new results in grey to indicate that they, as opposed to all the other numbers, have not appeared in the literature before. Also, to the

best of our knowledge, these results have not been compared simultaneously before.

We see in Table 2 that our test has the correct level and is quite powerful against all alternative specifications of conditional dependence, in the additive models 5 and 6, as well as the remaining examples 7 to 10, that are more multiplicative in nature.

When $n = 200$ we can include simulation results reported by Bouezmarni, Rombouts, and Taamouti (2012) for their conditional independence test based on measuring the Hellinger distance between copula density estimates (shortened BRT from the names of the authors) as well as the results reported by Bouezmarni and Taamouti (2014) on their test based on $L_2$ distances between estimated conditional distribution functions, which is abbreviated by BT. We include in Table 3* in the online supplement all results from these papers. The general picture is the same as in Table 2. In particular for the case with $c = 1.4$, our new test exhibits the best over-all performance among the examples listed, which, again to the best of our knowledge, include all such simulation results that have been published to date. It is seen that the standard linear Granger causality test (LIN) has a miserable performance in these examples.

Su and White (2014) define two extensions to a subset of the data generating processes defined in Table 1 in which the conditioning variable $\boldsymbol{X}^{(2)}$ is a vector with two and three components respectively. We refer to the online supplement for details regarding these extensions and numerical results. The new test for conditional independence based on the LGPC produces competitive results also in these cases. We point out again, however, that testing for conditional independence is merely one, albeit an important one, of several chief objectives for introducing the LGPC. Possible refinements of the LGPC testing is mentioned in the beginning of Section 5.

# 6 CONCLUSION AND OUTLOOK

The purpose of this paper is two-fold. First we define and develop the LGPC, a signed local measure of conditional dependence that has useful properties and is easy to interpret. It can measure both positive and negative conditional dependence. We then explore the possibility of using the LGPC in estimation of conditional dependence and indicate its potential in testing for conditional independence. The results are promising, and suggest that the newly developed tool provides powerful improvements to existing methods.

For 3 scalar variables $X_1, X_2, X_3$ a full trivariate analysis depending on all 3 coordinates can be undertaken as in Section 3.1.1. If any of the components $X_1, X_2$ or $X_3$ are vectors, a pairwise approximation and simplification has been described in Section 3.1.2. This can be likened to additive approximation in nonparametric regression.

Both the full trivariate approach and the pairwise simplification have been mainly analyzed at the $z$-scale using the transformed variables $\boldsymbol{Z} = \Phi^{-1}(F(\boldsymbol{X}))$. The corresponding LGPC $\alpha(\boldsymbol{z})$ is invariant to monotone transformations of the marginals. The LGPC $\alpha(\boldsymbol{x})$ on the $x$-scale can be obtained from $\alpha(\boldsymbol{z})$ by taking $\boldsymbol{x} = F^{-1}(\Phi(\boldsymbol{z}))$, and it is in general different from $\alpha(\boldsymbol{z})$ and is not invariant to monotone transformations of the marginals. An analogue is the different transformation properties of the Pearson correlation and the copula structure in describing joint dependence. For multivariate Gaussian variables, $\alpha(\boldsymbol{x}) \equiv \alpha(\boldsymbol{z}) \equiv \alpha$, which is the ordinary global partial correlation.

There is a potential for much further work. A next natural step may be structural equations as used in network analysis and Pearl-type causality (Pearl 2000), and ultimately further exploring the relationship or lack of relationship between that type of causality and Granger causality. All methods presented in this paper have been implemented in the R programming language (R Core Team 2020), and is available in the package **lg** (Otneim 2019). In addition, we provide code and data for easy replication of the results presented here in an online appendix.

# REFERENCES

Baba, Kunihiro, Ritei Shibata, and Masaaki Sibuya. 2004. "Partial Correlation and Conditional Correlation as Measures of Conditional Independence." *Australian & New Zealand Journal of Statistics* 46 (4): 657–64.

Baek, Ehung, and William Brock. 1992. "A General Test for Nonlinear Granger Causality: Bivariate Model." *Iowa State University and University of Wisconsin at Madison Working Paper.*

Bergsma, Wicher. 2010. "Nonparametric Testing of Conditional Independence by Means of the Partial Copula." *Unpublished Manuscript.* http://dx.doi.org/10.2139/ssrn.1702981.

Bouezmarni, Taoufik, Jeroen V K Rombouts, and Abderrahim Taamouti. 2012. "Nonparametric Copula-Based Test for Conditional Independence with Applications to Granger Causality." *Journal of Business and Economic Statistics* 30 (2): 275–87.

Bouezmarni, Taoufik, and Abderrahim Taamouti. 2014. "Nonparametric Tests for Conditional Independence Using Conditional Distributions." *Journal of Nonparametric Statistics* 26 (4): 697–719.

Cheng, Yu-Hsiang, and Tzee-Ming Huang. 2012. "A Conditional Independence Test for Dependent Data Based on Maximal Conditional Correlation." *Journal of Multivariate Analysis* 107: 210–26.

Delgado, Miguel A, and Wenceslao González Manteiga. 2001. "Significance Testing in Nonparametric Regression Based on the Bootstrap." *The Annals of Statistics* 29 (5): 1469–1507.

Francis, Bill B, Mbodja Mougoue, and Valentyn Panchenko. 2010. "Is There a Symmetric Nonlinear Causal Relationship Between Large and Small Firms?" *Journal of Empirical Finance* 17 (1): 23–38.

Granger, Clive WJ. 1969. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica*, 424–38.

———. 1980. "Testing for Causality: A Personal Viewpoint." *Journal of Economic Dynamics*

*and Control* 2: 329–52.

Hjort, Nils Lid, and MC Jones. 1996. "Locally Parametric Nonparametric Density Estimation." *The Annals of Statistics*, 1619–47.

Huang, Tzee-Ming. 2010. "Testing Conditional Independence Using Maximal Nonlinear Conditional Correlation." *The Annals of Statistics* 38 (4): 2047–91.

Jordanger, Lars Arne, and Dag Tjøstheim. 2020. "Nonlinear Spectral Analysis: A Local Gaussian Approach." *arXiv Preprint arXiv:1708.02166v4, Revised version submitted to the Journal of the American Statistical Association.*

Lacal, Virginia, and Dag Tjøstheim. 2017. "Local Gaussian Autocorrelation and Tests for Serial Independence." *Journal of Time Series Analysis* 38 (1): 51–71.

———. 2018. "Estimating and Testing Nonlinear Local Dependence Between Two Time Series." *Journal of Business & Economic Statistics*, 1–13.

Linton, Oliver, and Pedro Gozalo. 1997. "Conditional Independence Restrictions: Testing and Estimation." Unpublished manuscript.

———. 2014. "Testing Conditional Independence Restrictions." *Econometric Reviews* 33 (5-6): 523–52.

Otneim, Håkon. 2016. "Multivariate and Conditional Density Estimation Using Local Gaussian Approximations." PhD thesis, University of Bergen; The University of Bergen.

Otneim, Håkon. 2019. *lg: Locally Gaussian Distributions: Estimation and Methods.* https://CRAN.R-project.org/package=lg.

Otneim, Håkon, and Dag Tjøstheim. 2017. "The Locally Gaussian Density Estimator for Multivariate Data." *Statistics and Computing* 27 (6): 1595–1616.

———. 2018. "Conditional Density Estimation Using the Local Gaussian Correlation." *Statistics and Computing* 28 (2): 303–21.

Patra, Rohit K, Bodhisattva Sen, and Gábor J Székely. 2016. "On a Nonparametric Notion of Residual and Its Applications." *Statistics and Probability Letters* 109: 208–13.

Pearl, Judea. 2000. *Causality: Models, Reasoning and Inference.* Vol. 29. Springer.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Song, Kyungchul. 2009. "Testing Conditional Independence via Rosenblatt Transforms." *The Annals of Statistics* 37 (6B): 4011–45.

Su, Liangjun, and Halbert White. 2007. "A Consistent Characteristic Function-Based Test for Conditional Independence." *Journal of Econometrics* 141 (2): 807–34.

———. 2008. "A Nonparametric Hellinger Metric Test for Conditional Independence." *Econometric Theory* 24 (4): 829–64.

———. 2014. "Testing Conditional Independence via Empirical Likelihood." *Journal of Econometrics* 182 (1): 27–44.

Su, Liangjun, and Halbert L White. 2012. "Conditional Independence Specification Testing for Dependent Processes with Local Polynomial Quantile Regression." In *Essays in Honor of Jerry Hausman*, 355–434. Emerald Group Publishing Limited.

Teräsvirta, Timo, Dag Tjøstheim, and Clive W. J. Granger. 2010. *Modelling Nonlinear Economic Time Series.* Oxford University Press.

Tjøstheim, Dag, and Karl Ove Hufthammer. 2013. "Local Gaussian Correlation: A New Measure of Dependence." *Journal of Econometrics* 172 (1): 33–48.

Wang, Xia, and Yongmiao Hong. 2018. "Characteristic Function Based Testing for Conditional Independence: A Nonparametric Regression Approach." *Econometric Theory* 34 (4): 815–49.

Wang, Xueqin, Wenliang Pan, Wenhao Hu, Yuan Tian, and Heping Zhang. 2015. "Conditional Distance Correlation." *Journal of the American Statistical Association* 110 (512): 1726–34.

Zhang, Kun, Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. 2012. "Kernel-Based Conditional Independence Test and Application in Causal Discovery." *arXiv Preprint arXiv:1202.3775.*