

Online supplement accompanying ‘The Locally Gaussian Partial Correlation’

A THE MULTIVARIATE LGPC

A.1 Definition

The literature on nonlinear conditional dependence modeling is mostly concerned with the simple case where the vector of conditionally dependent variables consists of two scalar random variables, $\mathbf{Z}^{(1)} = (Z_1, Z_2)$, and this case is therefore the focus of our treatment of the LGPC in the main article. This means that we estimate a single local Gaussian partial correlation function $\hat{\alpha}(\mathbf{z})$. One of the most attractive properties of the LGPC, however, is the close parallel to the multivariate Gaussian distribution, and this is also the case when describing conditional dependence among more than two variables, which in our case leads to the estimation of a *matrix* of local Gaussian partial correlations. Besides needing a slightly more involved notation, this is a straightforward exercise in our framework as we will see below.

Let \mathbf{Z} be a vector of continuous random variables, that we split in two components $\mathbf{Z} = (\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)})$, but this time $\mathbf{Z}^{(1)} = (Z_1, \dots, Z_{d_1+d_2})$ is itself a vector with dimension $d_1 + d_2$ that is split in two groups; the d_1 -dimensional vector \mathbf{Z}_1 and the d_2 -dimensional vector \mathbf{Z}_2 . The vector $\mathbf{Z}^{(2)} = \mathbf{Z}_3$ is a vector of d_3 random variables. Denote by $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ the expectation vector and the covariance matrix of \mathbf{Z} , now partitioned according to our partition of $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3)$ as

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

where $\boldsymbol{\Sigma}_{11}$ is the covariance matrix of \mathbf{Z}_1 , $\boldsymbol{\Sigma}_{23}$ contains the covariances between \mathbf{Z}_2 and \mathbf{Z}_3 , and so forth, and where naturally $\boldsymbol{\Sigma}^T = \boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}_{12}^T = \boldsymbol{\Sigma}_{21}$, $\boldsymbol{\Sigma}_{13}^T = \boldsymbol{\Sigma}_{31}$, and $\boldsymbol{\Sigma}_{23}^T = \boldsymbol{\Sigma}_{32}$.

All of these parameters can be described and estimated locally in the exact same way as laid out in the main article, resulting from the locally approximating Gaussian joint density $\psi(\mathbf{z}, \mathbf{v})$ at the point $\mathbf{v} = \mathbf{z}$,

$$\psi(\mathbf{z}) = \frac{1}{(2\pi)^{(d_1+d_2+d_3)/2} |\boldsymbol{\Sigma}(\mathbf{z})|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{z} - \boldsymbol{\mu}(\mathbf{z}))^T \boldsymbol{\Sigma}^{-1}(\mathbf{z}) (\mathbf{z} - \boldsymbol{\mu}(\mathbf{z})) \right\} \quad (\text{A.1})$$

and the locally Gaussian *conditional* quantity

$$\begin{aligned}\psi(\mathbf{z}_1, \mathbf{z}_2 | \mathbf{Z}_3 = \mathbf{z}_3) &= \frac{1}{(2\pi)^{(d_1+d_2)/2} |\boldsymbol{\Sigma}_{12|3}(\mathbf{z})|^{1/2}} \\ &\times \exp \left\{ -\frac{1}{2} (\mathbf{z} - \boldsymbol{\mu}_{12|3}(\mathbf{z}))^T \boldsymbol{\Sigma}_{12|3}^{-1}(\mathbf{z}) (\mathbf{z} - \boldsymbol{\mu}_{12|3}(\mathbf{z})) \right\}.\end{aligned}$$

In order to simplify notation, in the following we will drop the z -dependence in the parameters. In order to define $\boldsymbol{\mu}_{12|3}$ and $\boldsymbol{\Sigma}_{12|3}$ let $\boldsymbol{\Sigma}^{11}$, $\boldsymbol{\Sigma}^{12}$, $\boldsymbol{\Sigma}^{21}$ and $\boldsymbol{\Sigma}^{22}$ be matrices of dimension $(d_1 + d_2) \times (d_1 + d_2)$, $(d_1 + d_2) \times d_3$, $d_3 \times (d_1 + d_2)$ and $d_3 \times d_3$ respectively given by

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

such that

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}^{11} & \boldsymbol{\Sigma}^{12} \\ \boldsymbol{\Sigma}^{21} & \boldsymbol{\Sigma}^{22} \end{pmatrix},$$

It follows from standard arguments (cf. Otneim and Tjøstheim (2018)) that

$$\boldsymbol{\mu}_{12|3} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} + \boldsymbol{\Sigma}^{12} (\boldsymbol{\Sigma}^{22})^{-1} (\mathbf{z}^{(2)} - \boldsymbol{\mu}_3) = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} + \begin{pmatrix} \boldsymbol{\Sigma}_{13} \boldsymbol{\Sigma}_{33}^{-1} (\mathbf{z}^{(2)} - \boldsymbol{\mu}_3) \\ \boldsymbol{\Sigma}_{23} \boldsymbol{\Sigma}_{33}^{-1} (\mathbf{z}^{(2)} - \boldsymbol{\mu}_3) \end{pmatrix} \quad (\text{A.2})$$

and

$$\begin{aligned}\boldsymbol{\Sigma}_{12|3} &= \boldsymbol{\Sigma}^{11} - \boldsymbol{\Sigma}^{12} (\boldsymbol{\Sigma}^{22})^{-1} \boldsymbol{\Sigma}^{21} = \boldsymbol{\Sigma}^{11} - \begin{pmatrix} \boldsymbol{\Sigma}_{13} \\ \boldsymbol{\Sigma}_{23} \end{pmatrix} \boldsymbol{\Sigma}_{33}^{-1} \begin{pmatrix} \boldsymbol{\Sigma}_{31} & \boldsymbol{\Sigma}_{32} \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{13} \boldsymbol{\Sigma}_{33}^{-1} \boldsymbol{\Sigma}_{31} & \boldsymbol{\Sigma}_{12} - \boldsymbol{\Sigma}_{13} \boldsymbol{\Sigma}_{33}^{-1} \boldsymbol{\Sigma}_{32} \\ \boldsymbol{\Sigma}_{21} - \boldsymbol{\Sigma}_{23} \boldsymbol{\Sigma}_{33}^{-1} \boldsymbol{\Sigma}_{31} & \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{23} \boldsymbol{\Sigma}_{33}^{-1} \boldsymbol{\Sigma}_{32} \end{pmatrix}.\end{aligned} \quad (\text{A.3})$$

We proceed now to establish the notion of conditional independence between \mathbf{Z}_1 and \mathbf{Z}_2 using the LGPC in the multivariate case. The conditional densities of $\mathbf{Z}_1 | \mathbf{Z}_3$ and $\mathbf{Z}_2 | \mathbf{Z}_3$ in terms of the local Gaussian parameters at the point $(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3)$ are given by

$$\begin{aligned}f(\mathbf{z}_1 | \mathbf{z}_3) &= \frac{1}{(2\pi)^{d_1/2} |\boldsymbol{\Sigma}_{1|3}|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{z}_1 - \boldsymbol{\mu}_{1|3})^T \boldsymbol{\Sigma}_{1|3}^{-1} (\mathbf{z}_1 - \boldsymbol{\mu}_{1|3}) \right\}, \\ f(\mathbf{z}_2 | \mathbf{z}_3) &= \frac{1}{(2\pi)^{d_2/2} |\boldsymbol{\Sigma}_{2|3}|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{z}_2 - \boldsymbol{\mu}_{2|3})^T \boldsymbol{\Sigma}_{2|3}^{-1} (\mathbf{z}_2 - \boldsymbol{\mu}_{2|3}) \right\}\end{aligned}$$

where

$$\boldsymbol{\mu}_{1|3} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}^{13} (\boldsymbol{\Sigma}^{33})^{-1} (\mathbf{z}_3 - \boldsymbol{\mu}_3), \quad \boldsymbol{\mu}_{2|3} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}^{23} (\boldsymbol{\Sigma}^{33})^{-1} (\mathbf{z}_3 - \boldsymbol{\mu}_3) \quad (\text{A.4})$$

and

$$\boldsymbol{\Sigma}_{1|3} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{13} \boldsymbol{\Sigma}_{33}^{-1} \boldsymbol{\Sigma}_{31}, \quad \boldsymbol{\Sigma}_{2|3} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{23} \boldsymbol{\Sigma}_{33}^{-1} \boldsymbol{\Sigma}_{32}. \quad (\text{A.5})$$

Under conditional independence between \mathbf{Z}_1 and \mathbf{Z}_2 given \mathbf{Z}_3 , the joint conditional density of $(\mathbf{Z}_1, \mathbf{Z}_2 | \mathbf{Z}_3)$ at the point $(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3)$ is given by

$$\begin{aligned} f(\mathbf{z}_1, \mathbf{z}_2 | \mathbf{z}_3) &= f(\mathbf{z}_1 | \mathbf{z}_3) f(\mathbf{z}_2 | \mathbf{z}_3) \\ &= \frac{1}{(2\pi)^{(d_1+d_2)/2} |\Sigma_{1|3}|^{1/2} |\Sigma_{2|3}|^{1/2}} \\ &\quad \times \exp \left\{ -\frac{1}{2} \left[(\mathbf{z}_1 - \boldsymbol{\mu}_{1|3})^T \Sigma_{1|3}^{-1} (\mathbf{z}_1 - \boldsymbol{\mu}_{1|3}) + \right. \right. \\ &\quad \left. \left. (\mathbf{z}_2 - \boldsymbol{\mu}_{2|3})^T \Sigma_{2|3}^{-1} (\mathbf{z}_2 - \boldsymbol{\mu}_{2|3}) \right] \right\}. \end{aligned}$$

Comparing expressions (A.1) – (A.5) we see that we have conditional independence if and only if the matrix $\Sigma_{12|3}$ in (A.3) is block diagonal, i.e. if and only if

$$\Sigma_{12} - \Sigma_{13} \Sigma_{33}^{-1} \Sigma_{32} = \Sigma_{21} - \Sigma_{23} \Sigma_{33}^{-1} \Sigma_{31} = \mathbf{0}.$$

A.2 Some special cases

Let $\mathbf{D} = \text{diag} \{ \Sigma_{12|3} \}$. The local Gaussian partial correlation matrix can be expressed as the lower left hand block matrix of $\mathbf{D}^{-1/2} \Sigma_{12|3} \mathbf{D}^{-1/2}$, and so the LGPC matrix is thus given as the lower left hand block of the matrix

$$\mathbf{D}^{-1/2} \Sigma_{12|3} \mathbf{D}^{-1/2} = \mathbf{D}^{-1/2} \begin{pmatrix} \Sigma_{11} - \Sigma_{13} \Sigma_{33}^{-1} \Sigma_{31} & \Sigma_{12} - \Sigma_{13} \Sigma_{33}^{-1} \Sigma_{32} \\ \Sigma_{21} - \Sigma_{23} \Sigma_{33}^{-1} \Sigma_{31} & \Sigma_{22} - \Sigma_{23} \Sigma_{33}^{-1} \Sigma_{32} \end{pmatrix} \mathbf{D}^{-1/2}. \quad (\text{A.6})$$

Let us consider this expression for a few special cases.

- a) Let $\mathbf{d}_1 = \mathbf{d}_2 = \mathbf{d}_3 = \mathbf{1}$. Consider the partial covariance matrix, where we again drop the z -dependence in all parameters:

$$\Sigma_{12|3} = \Sigma^{11} - \Sigma^{12} (\Sigma^{22})^{-1} \Sigma^{21}.$$

Here, since all variables are normalized with variance one, we have that

$$\Sigma = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{21} & 1 & \rho_{23} \\ \rho_{31} & \rho_{32} & 1 \end{pmatrix},$$

and

$$\Sigma^{11} = \begin{pmatrix} 1 & \rho_{12} \\ \rho_{21} & 1 \end{pmatrix}, \quad \Sigma^{12} = \begin{pmatrix} \rho_{13} \\ \rho_{23} \end{pmatrix}, \quad (\Sigma^{22})^{-1} = 1, \quad \Sigma^{21} = \begin{pmatrix} \rho_{31} & \rho_{32} \end{pmatrix}.$$

Hence,

$$\begin{aligned} \Sigma^{11} - \Sigma^{12} (\Sigma^{22})^{-1} \Sigma^{21} &= \begin{pmatrix} 1 & \rho_{12} \\ \rho_{21} & 1 \end{pmatrix} - \begin{pmatrix} \rho_{13} \\ \rho_{23} \end{pmatrix} \begin{pmatrix} \rho_{31} & \rho_{32} \end{pmatrix} \\ &= \begin{pmatrix} 1 - \rho_{13}^2 & \rho_{12} - \rho_{13}\rho_{23} \\ \rho_{12} - \rho_{13}\rho_{23} & 1 - \rho_{23}^2 \end{pmatrix}. \end{aligned}$$

This is a covariance matrix, and not a correlation matrix, so normalizing with standard deviations results in the partial correlation matrix

$$\begin{pmatrix} 1 & \frac{\rho_{12}-\rho_{13}\rho_{23}}{\sqrt{1-\rho_{13}^2}\sqrt{1-\rho_{23}^2}} \\ \frac{\rho_{12}-\rho_{13}\rho_{23}}{\sqrt{1-\rho_{13}^2}\sqrt{1-\rho_{23}^2}} & 1 \end{pmatrix}. \quad (\text{A.7})$$

This result can also be obtained by direct application of (A.6):

$$\begin{pmatrix} 1-\rho_{13}^2 & 0 \\ 0 & 1-\rho_{23}^2 \end{pmatrix}^{-1/2} \begin{pmatrix} 1-\rho_{13}^2 & \rho_{12}-\rho_{13}\rho_{23} \\ \rho_{12}-\rho_{13}\rho_{23} & 1-\rho_{23}^2 \end{pmatrix} \begin{pmatrix} 1-\rho_{13}^2 & 0 \\ 0 & 1-\rho_{23}^2 \end{pmatrix}^{-1/2}.$$

b) Let $d_3 = 1$ and let d_1 and d_2 be arbitrary. In this case

$$\Sigma^{11} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

where

$$\begin{aligned} \Sigma_{11} &= \{\rho_{ij}^{1,1}\}, i = 1, \dots, d_1; j = 1, \dots, d_1, & \Sigma_{12} &= \{\rho_{ij}^{1,2}\}, i = 1, \dots, d_1; j = 1, \dots, d_2 \\ \Sigma_{21} &= \{\rho_{ij}^{2,1}\}, i = 1, \dots, d_2; j = 1, \dots, d_1, & \Sigma_{22} &= \{\rho_{ij}^{2,2}\}, i = 1, \dots, d_2; j = 1, \dots, d_2. \end{aligned}$$

Since $\Sigma^{22} = \Sigma_{33} = 1$,

$$\Sigma^{12} (\Sigma^{22})^{-1} \Sigma^{21} = \Sigma^{12} \Sigma^{21} = \begin{pmatrix} \Sigma_{13} \\ \Sigma_{23} \end{pmatrix} (\Sigma_{31} \Sigma_{32}),$$

where

$$\begin{aligned} \Sigma_{13} &= \{\rho_{i3}^{1,3}\}, i = 1, \dots, d_1; & \Sigma_{23} &= \{\rho_{j3}^{2,3}\}, j = 1, \dots, d_2 \\ \Sigma_{31} &= \{\rho_{3i}^{3,1}\}, i = 1, \dots, d_1; & \Sigma_{32} &= \{\rho_{3j}^{3,2}\}, j = 1, \dots, d_2. \end{aligned}$$

Furthermore,

$$\Sigma_{12|3} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} - \begin{pmatrix} \Sigma_{13} \\ \Sigma_{23} \end{pmatrix} (\Sigma_{31} \Sigma_{32}).$$

The diagonal of Σ_{11} and Σ_{22} are populated with ones, whereas, using symmetry of correlations, $\begin{pmatrix} \Sigma_{13} & \Sigma_{23} \end{pmatrix}^T \begin{pmatrix} \Sigma_{31} & \Sigma_{32} \end{pmatrix}$ has $(\rho_{i3}^{1,2})^2$, $i = 1, \dots, d_1$ and $(\rho_{3j}^{3,2})^2$, $j = 1, \dots, d_2$ along the diagonal, such that \mathbf{D} is the $(d_1 + d_2) \times (d_1 + d_2)$ diagonal matrix with the concatenated sequence $1 - (\rho_{i3}^{1,2})^2$, $i = 1, \dots, d_1$; $1 - (\rho_{3j}^{3,2})^2$, $j = 1, \dots, d_2$ along the diagonal.

Pre- and post multiplying $\Sigma_{12|3}$ by $\mathbf{D}^{-1/2}$ converts $\Sigma_{12|3}$ into a correlation matrix. All of the off-diagonal elements in this matrix has the same form as the off-diagonal

elements in (A.7) so that

$$\rho(Z_{1,i}, Z_{1,j}|Z_3) = \frac{\rho(Z_{1,i}, Z_{1,j}) - \rho(Z_{1,i}, Z_3)\rho(Z_{1,j}, Z_3)}{\sqrt{1 - \rho^2(Z_{1,i}, Z_3)}\sqrt{1 - \rho^2(Z_{1,j}, Z_3)}}, \quad i, j = 1, \dots, d_1, \quad (\text{A.8})$$

$$\rho(Z_{2,i}, Z_{2,j}|Z_3) = \frac{\rho(Z_{2,i}, Z_{2,j}) - \rho(Z_{2,i}, Z_3)\rho(Z_{2,j}, Z_3)}{\sqrt{1 - \rho^2(Z_{2,i}, Z_3)}\sqrt{1 - \rho^2(Z_{2,j}, Z_3)}}, \quad i, j = 1, \dots, d_2; \quad (\text{A.9})$$

$$\rho(Z_{1,i}, Z_{2,j}|Z_3) = \frac{\rho(Z_{1,i}, Z_{2,j}) - \rho(Z_{1,i}, Z_3)\rho(Z_{2,j}, Z_3)}{\sqrt{1 - \rho^2(Z_{1,i}, Z_3)}\sqrt{1 - \rho^2(Z_{2,j}, Z_3)}}, \quad i = 1, \dots, d_1; j = 1, \dots, d_2. \quad (\text{A.10})$$

Testing for conditional independence between \mathbf{Z}_1 and \mathbf{Z}_2 given \mathbf{Z}_3 is in our context the same as testing for zero local partial correlation between all pairs of variables, a test for which is presented in the main article and discussed in technical detail later in this supplement. One can also, based on (A.8) and (A.9), test for conditional independence between combinations of components given \mathbf{Z}_3 . Equations (A.8)–(A.10) above give explicit expressions for the conditional correlation matrix of \mathbf{Z}_1 and \mathbf{Z}_2 given \mathbf{Z}_3 , where estimated values give a measure of local conditional dependence between the components.

- c) **Let $d_1 = d_2 = 1$ and let d_3 be arbitrary.** This is the case usually treated in the literature, also the focus of the discussions in the main article. In this case

$$\Sigma^{11} = \begin{pmatrix} 1 & \rho_{12} \\ \rho_{21} & 1 \end{pmatrix}, \Sigma_{13} = \{\rho_{1i}^{1,3}\}, i = 1, \dots, d_3; \Sigma_{23} = \{\rho_{2i}^{2,3}\}; i = 1, \dots, d_3,$$

hence $(\Sigma_{13} \quad \Sigma_{23})^T$ is of dimension $2 \times d_3$, whereas $(\Sigma_{31} \quad \Sigma_{32})$ is of dimension $d_3 \times 2$ such that

$$\begin{pmatrix} \Sigma_{13} \\ \Sigma_{23} \end{pmatrix} \Sigma_{33}^{-1} (\Sigma_{31} \quad \Sigma_{32})$$

is of dimension 2×2 . However, as is immediately seen, no formulas like (A.8)–(A.10) will appear; they will be much more complicated and include sums and products of various correlations.

- d) **Let d_1, d_2 and d_3 be arbitrary.** This is the general case, and it seems impossible to obtain simple explicit formulas for the local Gaussian partial correlations, but the general fact still holds of course: The local partial correlation matrix of \mathbf{Z}_1 and \mathbf{Z}_2 given \mathbf{Z}_3 can be identified by the lower left hand block matrix of $\mathbf{D}^{-1/2} \Sigma_{12|3} \mathbf{D}^{-1/2}$, and can be estimated using pairwise local correlations in (A.6).

A.3 Estimation and estimation theory in more general situations

Sections B – D of this supplement deal with the results stated in Section 3 of the main article where X_1 and X_2 are scalars. The present subsection looks briefly at more general situations.

Recall that the matrix D defined in the beginning of Section A.2 is given by $D = \text{diag}(\mathbf{\Sigma}_{12|3})$. Let us block divide the $(d_1 + d_2) \times (d_1 + d_2)$ diagonal matrix D into

$$D = \begin{pmatrix} D_{11} & \mathbf{0} \\ \mathbf{0} & D_{22} \end{pmatrix}.$$

The estimated partial correlation matrix between \mathbf{X}_1 and \mathbf{X}_2 given \mathbf{X}_3 is then given as

$$\widehat{D}_{11}^{-1/2} \left(\widehat{\mathbf{\Sigma}}_{21} - \widehat{\mathbf{\Sigma}}_{23} \widehat{\mathbf{\Sigma}}_{33}^{-1} \widehat{\mathbf{\Sigma}}_{31} \right) \widehat{D}_{22}^{-1/2} \stackrel{\text{def}}{=} g(\widehat{\mathbf{\Sigma}}). \quad (\text{A.11})$$

We present the asymptotic theory for the local likelihood estimator of the local partial correlations in two situations in the main article with details of proof in Sections B – D:

- (i) If $d_1 = d_2 = d_3 = 1$, then we can estimate the full trivariate local likelihood estimate, meaning that all the local correlations depend on all three coordinates $z = (z_1, z_2, z_3)$. The asymptotic theory for this estimator is covered by Theorem 3.1 in the main article.
- (ii) If $d_1 = d_2 = 1$ and $d_3 > 1$, we estimate all correlations in a pairwise manner as laid out in the main article. The limiting distribution of the (scalar) local Gaussian partial correlation is in general given by

$$\sqrt{nb_n^2} (\widehat{\alpha}(z)) - \alpha(z)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nabla g(\boldsymbol{\rho})^T \mathbf{\Lambda} \nabla g(\boldsymbol{\rho})), \quad (\text{A.12})$$

see eq. (22) in the main article, and see Theorem 2 in Otneim and Tjøstheim (2018) with $\mathbf{\Sigma} = \mathbf{\Lambda}$ in that theorem. The computation of ∇g is outlined in Section C.

These results can be extended to the multivariate cases that we treat in this supplement in Section A.2 as well. The assumptions remain identical to the assumptions in Theorem 3.1 in the main article and Theorem 3 in Otneim and Tjøstheim (2018).

- (iii) If d_1 and d_2 are arbitrary, and $d_3 = 1$, the local partial correlations take the form (A.10). The limiting distribution of $\widehat{\rho}(Z_{1i}, Z_{2j} | Z_3)$ follows from the delta rule applied to the joint distribution of $\{\widehat{\rho}(Z_{1i}, Z_{2j}), \widehat{\rho}(Z_{1i}, Z_3), \widehat{\rho}(Z_{2j}, Z_3)\}$, the joint distribution of which again follows directly from Theorem 2 in Otneim and Tjøstheim (2018), who also note that the asymptotic correlation between two local correlation estimates is of smaller order if the two index pairs are distinct. The computation of ∇g is fairly straightforward in this case, by taking the appropriate derivatives in (A.10).
- (iv) If d_1, d_2 and d_3 are all arbitrary, we have the general case, where, in principle, for each $i = 1, \dots, d_1$ and $j = 1, \dots, d_2$, the local partial correlation $\rho(Z_{1i}, Z_{2j} | \mathbf{Z}_3)$ can be expressed as a deterministic function g_{ij} of a local correlation vector $\boldsymbol{\rho}$ using equation (A.11), the asymptotic normality for which is proved as before using the delta rule and Theorem 2 of Otneim and Tjøstheim (2018). It is, however, difficult to obtain explicit expressions of the asymptotic covariance matrix in this case.

The joint asymptotic normality of the pairwise partial correlation estimates can be proved using the Cramer-Wold device. But, again, obtaining an expression for the asymptotic covariance matrix is hard in the general case.

Observe, finally, that a more general pairwise scheme can be devised, as treated in detail by Lacal and Tjøstheim (2017), Lacal and Tjøstheim (2019) and Jordanger and Tjøstheim (2020):

- (v) In this scheme, we do not use a pairwise set-up for the local correlation only (the one-parameter case), but for all parameters in a bivariate normal distribution; the correlation $\rho(x_1, x_2)$, the expectations $\mu_1(x_1, x_2), \mu_2(x_1, x_2)$ and the variances $\sigma_1(x_1, x_2), \sigma_2(x_1, x_2)$ (the five-parameter case). This setup can also be simplified and fitted to a p -variate distribution by estimating $\mu_i = \mu_i(x_i)$ and $\sigma_i = \sigma_i(x_i)$, $i = 1, \dots, p$, marginally. Tjøstheim and Hufthammer (2013) give the asymptotic theory in this case.

Below is a brief summary of the relevant local likelihood theory for the general pairwise case of (ii)–(iv). We also state a general consistency result in Theorem A.1 at the end of this section. Utilizing Theorem 3.2 in the main article we replace the pseudo normals $\widehat{\mathbf{Z}}_t$ of equation (12) of the main paper by vectors \mathbf{Z}_t having standard normal marginals.

The pairwise log-likelihood in cases (ii) – (iv) can then be written as

$$\begin{aligned} L_{n,b}(\mathbf{Z}_{in}, \mathbf{Z}_{jn}; \rho_b(z_i, z_j)) &= \frac{1}{n} \sum_{t=1}^n K_b(Z_{i,t} - z_i, Z_{j,t} - z_j) \log(\psi(Z_{i,t}, Z_{j,t}; \rho_b(z_i, z_j))) \\ &\quad - \int K_b(v - z_i, w - z_j) \psi(v, w; \rho_b(z_i, z_j)) \, dv \, dw, \end{aligned} \quad (\text{A.13})$$

where $\mathbf{Z}_{in} = (Z_{i,1}, Z_{i,2}, \dots, Z_{i,n})$ and $\mathbf{Z}_{jn} = (Z_{j,1}, Z_{j,2}, \dots, Z_{j,n})$ are the \mathbf{Z} -vectors for components i and j respectively, n is the number of observations, $z_i, z_j, v, w \in \mathbb{R}$, $\rho_b = \rho_{ij,b}(z_i, z_j)$ is the local Gaussian correlation parameter, $K_b(v - z_i, w - z_j) = (b_1 b_2)^{-1} K(b_1^{-1}(v - z_i)) K(b_2^{-1}(w - z_j))$ is a product kernel with bandwidth $b = b_n = (b_1, b_2) = (b_{1n}, b_{2n})$, and K is a non-negative kernel function with compact support such that $\int K(v, w) = 1$. For a fixed bandwidth b , $\rho_{0,b}$ is the true local parameter, that is, the minimizer of the local penalty function

$$\int K_b(v - z_i, w - z_j) [\psi(v, w; \rho_b(z_i, z_j)) - \log(\psi(v, w; \rho_b(z_i, z_j))) f(v, w)] \, dv \, dw, \quad (\text{A.14})$$

where f is the density function of (Z_i, Z_j) . Note in particular that $\rho_{0,b} \rightarrow \rho_0$ as $b \rightarrow 0$, which is the true local correlation parameter. The reader may consult Tjøstheim and Hufthammer (2013) for more details and regularity conditions in this regard. The corresponding estimate $\hat{\rho} = \rho_{n,b}(z_i, z_j)$ is obtained by maximizing the log-likelihood in (A.13). Note that, writing ρ for ρ_b ,

$$L_{n,b}(\rho) = \frac{1}{n} \sum_{t=1}^n L^{(b)}(Z_{i,t}, Z_{j,t}; \rho), \quad (\text{A.15})$$

where

$$L^{(b)}(Z_{i,t}, Z_{j,t}; \rho) = K_b(Z_{i,t} - z_i, Z_{j,t} - z_j) \log(\psi(Z_{i,t}, Z_{j,t}; \rho_b(z_i, z_j))) \\ - \int K_b(v - z_i, w - z_j) \psi(v, w; \rho_b(z_i, z_j)) \, dv \, dw. \quad (\text{A.16})$$

We must next proceed by introducing some new concepts before stating the equivalents of assumptions A1 – A2 by Lacal and Tjøstheim (2019). These assumptions are similar but distinct from those used in Theorem 3.1 in the main article and Theorem 2 in Otneim and Tjøstheim (2018). They will also be used to prove the consistency result in Theorem A.1 at the end of this section, and they will be part of the assumptions used to prove the validity of the bootstrap in Section E.

For a given function g , the process $\{g(Z_{i,t}, Z_{j,t}; \rho)\}$ is said to be r -dominated on Θ uniformly in t , where $t = 1, \dots, n$, if there exists a function $D_t : \mathbb{R} \rightarrow \mathbb{R}$ such that $|g(Z_{i,t}, Z_{j,t}; \rho)| \leq D_t$ for all $\rho \in \Theta$, $t = 1, \dots, n$ and D_t is \mathcal{B} -measurable such that for some $\Delta > 0$, $\|D_t\|_r \stackrel{\text{def}}{=} (\int |D_t(u)|^r \, du)^{1/r} \leq \Delta < \infty$ for all $t = 1, \dots, n$, where \mathcal{B} is the Borel σ -algebra. In this particular case the location (z_i, z_j) is fixed, and $\Theta = [-1, 1]$. Moreover, the process $\{g(Z_{i,t}, Z_{j,t}; \rho)\}$ is said to be Lipschitz continuous on Θ if

$$|\{g(Z_{i,t}, Z_{j,t}; \rho)\} - \{g(Z_{i,t}, Z_{j,t}; \rho')\}| \leq C_t |\rho - \rho'|$$

almost surely in \mathbb{P} for all $\rho, \rho' \in \Theta$ and with a sufficiently large constant M such that $C_t \leq M$. Here \mathbb{P} is the probability measure on the probability space in which $\{Z_{i,t}, Z_{j,t}\}$ is defined. We can now state conditions A1 – A2 which are basic assumptions for proving consistency and asymptotic normality of $\rho_{n,b}$:

A1 $(\Omega, \mathcal{F}, \mathbb{P})$ is a complete probability space on which is defined a sequence of random functions $\{L_{n,b} : \Omega \rightarrow \overline{\mathbb{R}}\}$, where $\overline{\mathbb{R}}$ is the extended real line. The parameter set Θ is compact (trivially fulfilled in our case because $\Theta = [-1, 1]$). The random process $\{Z_{i,t}, Z_{j,t}\}$ is \mathcal{F}_t -measurable, stationary and α -mixing with mixing coefficients $\alpha_k = O(\alpha^k)$, $\alpha \in (0, 1)$ for all t , and with a bivariate density having support in all of \mathbb{R}^2 . Here \mathcal{F}_t is the σ -algebra generated by $\{Z_{i,s}, Z_{j,s}, s \leq t\}$.

A2 The parameter value $\rho_{0,b}$ is the unique maximizer of $E(L_{n,b}(\rho)) : \Theta \rightarrow \overline{\mathbb{R}}$, with $E(L_{n,b}(\rho))$ defined in (A.14).

The conditions A1– A2 are quite mild and discussed by Lacal and Tjøstheim (2019). Note that they have been formulated for a scalar parameter $\rho = \rho_{ij}$. To have them fulfilled for all possible combinations of i, j , the assumption A1, for example can be easily formulated for the $\{\mathbf{Z}_t\}$ process. Consistency for the local parameter estimates is required for consistency of the bootstrap:

Theorem A.1. *Let assumptions A1 and A2 hold, and let $\{L^{(b)}(Z_{i,t}, Z_{j,t}; \rho)\}, \rho \in \Theta$ be Lipschitz continuous on Θ and 4 -dominated on Θ uniformly in t and n . Moreover, assume that as $b = b_n \rightarrow 0$, $\sigma_b^2 \stackrel{\text{def}}{=} \text{Var}(L^{(b)}(Z_{i,t}, Z_{j,t}; \rho)) \rightarrow \sigma^2 < \infty$ and $\mu_b \stackrel{\text{def}}{=} E(L^{(b)}(Z_{i,t}, Z_{j,t}; \rho)) \rightarrow \mu < \infty$. Then $\rho_{n,b} - \rho_{0,b} \xrightarrow{\mathbb{P}} 0$ for a fixed b and $\rho_{n,b} - \rho_0 \xrightarrow{\mathbb{P}} 0$ as $b \rightarrow 0$.*

The proof of Theorem A.1 is mostly identical to the proof of Theorem 2.1 of Lacial and Tjøstheim (2019) with trivial modifications to take into account the scalar nature of the parameter ρ_{ij} in the present case. A distributional limit theorem analogous to Theorem 2.2 in Lacial and Tjøstheim (2019) can also be rephrased and proved as in that paper, but we refrain from doing this given the distributional results already stated in Section 3 of the main part of the paper.

B DETAILS REGARDING THE PROOF OF THEOREM 3.1

For fixed bandwidths we have from standard arguments, such as those provided by Hjort and Jones (1996), that

$$\sqrt{nb^3}(\hat{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_0) \xrightarrow{d} \mathcal{N}(0, \mathbf{J}_b^{-1} \mathbf{M}_b (\mathbf{J}_b^{-1})^T), \quad (\text{B.1})$$

which corresponds to the usual rate for nonparametric density estimation. As $\mathbf{b} \rightarrow 0$, however, we must take extra care when considering the asymptotic covariance matrix. Tjøstheim and Hufthammer (2013), Section 4, write the Taylor expansions of \mathbf{M}_b and \mathbf{J}_b , now being functionals of three variables (x_1, x_2, x_3) or (z_1, z_2, z_3) , as

$$\mathbf{M}_b = I_M + II_M + o(b^2) \quad \text{and} \quad \mathbf{J}_b = I_J + II_J + o(b^2),$$

and consider each term. For example, we write their first term in the expansion of \mathbf{M}_b in the three-variate case as follows:

$$I_M \sim \int K^2(w_1, w_2, w_3) A b_w b_w^T A^T f(\mathbf{z} + b_1 w_1 + b_2 w_2 + b_3 w_3) dw_1 dw_2 dw_3,$$

where K is a product kernel, and, in our case, b_w is the vector defined by $b_w^T = (1 \ b_1 w_1 \ b_2 w_2 \ b_3 w_3)$, and A is the 3×4 matrix $A = \begin{pmatrix} v & v_{z_1} & v_{z_2} & v_{z_3} \end{pmatrix}$ with $\mathbf{v}^T = (v_1 \ v_2 \ v_3)$, $v_i = \partial L_n / \partial \rho_i$, where the index $i = 1, 2, 3$ represents the three local correlations, and $v_{z_i} = \partial \mathbf{v} / \partial z_i$. In the next step, they Taylor expand f and compute the matrix $\int K^2(w_1, w_2, w_3) b_w b_w^T dw_1 dw_2 dw_3$, which, by omitting all constant factors and exploiting that $\int w^k K^2(w) dw = 0$ for $k = 1, 3$, in our case becomes the diagonal matrix

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & b_1^2 w_1^2 & 0 & 0 \\ 0 & 0 & b_2^2 w_2^2 & 0 \\ 0 & 0 & 0 & b_3^2 w_3^2 \end{pmatrix}.$$

The second term defining \mathbf{M}_b has smaller order than I_M as $b \rightarrow 0$, while the first term of \mathbf{J}_b can be treated exactly as I_M above resulting in a matrix of order b^2 .

The second term in \mathbf{J}_b can similarly be written as $II_J = B(\psi(\mathbf{z}, \boldsymbol{\rho}_b(\mathbf{z})) - f(\mathbf{z})) + o(b^2)$, where $B = \nabla \mathbf{v}(\mathbf{z}, \boldsymbol{\rho}_b(\mathbf{z}))$ are second derivatives of the local likelihood function with respect to the local correlations, and $\psi(\mathbf{z}, \boldsymbol{\rho}_b(\mathbf{z})) - f(\mathbf{z}) = O(b^2)$ according to Hjort and Jones (1996). Hence, it follows that the leading term of $\mathbf{J}_b^{-1} \mathbf{M}_b (\mathbf{J}_b^{-1})^T$ is $O(b^{-2})$, which again means that the convergence rate in eq. (B.1) above must be modified accordingly in order to balance the

convergence in eq. (20) in the main article. Extracting an analytic expression for the leading term of the asymptotic covariance matrix is, as mentioned by Tjøstheim and Hufthammer (2013), possible by means of symbolic manipulation software, but the expression is very complicated and of little practical use. See also the corresponding proof of Theorem 3 in Tjøstheim and Hufthammer (2013).

C THE ASYMPTOTIC VARIANCE OF THE LGPC

The limiting distribution of the LGPC is given in eq. (22) in the main article. In this section, we will present the basic steps needed to work out the value of $\nabla g(\boldsymbol{\rho})$ in the simplest case where $d_1 = d_2 = 1$. In the more general situation it is not feasible to obtain analytic expressions for the asymptotic variance, and the bootstrap should be used. For the sake of this particular argument, we may simplify notation quite a bit, leaving out what we do not need: We drop the \mathbf{z} -dependence, and simply write $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{12|3}(\mathbf{z})$ for the 2×2 local partial covariance matrix between Z_1 and Z_2 given (Z_3, \dots, Z_p) . The vector of local correlations $\boldsymbol{\rho}$ is indexed by k , which in turn means that k corresponds to a specific *pair* (i, j) : $\rho_k = \rho_{ij}$, with $\rho_{ij} = \rho_{ji}$. Let $\boldsymbol{\Sigma}^{(k)}$ be the element-wise matrix of partial derivatives

$$\boldsymbol{\Sigma}^{(k)} = \frac{\partial \boldsymbol{\Sigma}}{\partial \rho_k} = \frac{\partial \boldsymbol{\Sigma}}{\partial \rho_{ij}},$$

which means that element k in the gradient $\nabla g(\boldsymbol{\rho})$ is given by

$$\{\nabla g(\boldsymbol{\rho})\}_k = \frac{\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{22}\boldsymbol{\Sigma}_{12}^{(k)} - \boldsymbol{\Sigma}_{12}\left(\boldsymbol{\Sigma}_{11}^{(k)}\boldsymbol{\Sigma}_{22} + \boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{22}^{(k)}\right)}{(\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{22})^2}, \quad (\text{C.1})$$

where the double subscript to $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^{(k)}$ here means matrix elements. In order to avoid confusion with the double subscripts defining matrix *partitions* in the preceding sections, we re-label the matrix blocks \mathbf{R}^{11} , \mathbf{R}^{12} and \mathbf{R}^{22} to \mathbf{R}_1 , \mathbf{R}_2 and \mathbf{R}_3 respectively, with $\mathbf{R}^{21} = \mathbf{R}_2^T$. Following basic differentiation rules for matrices (see e.g. Van den Bos (2007)), we have that, element-wise,

$$\frac{\partial \left(\mathbf{R}_2\mathbf{R}_3^{-1}\mathbf{R}_2^T\right)}{\partial \rho_k} = \frac{\partial \mathbf{R}_2}{\partial \rho_k}\mathbf{R}_3^{-1}\mathbf{R}_2^T - \mathbf{R}_2\mathbf{R}_3^{-1}\frac{\partial \mathbf{R}_3}{\partial \rho_k}\mathbf{R}_3^{-1}\mathbf{R}_2^T + \mathbf{R}_2\mathbf{R}_3^{-1}\frac{\partial \mathbf{R}_2^T}{\partial \rho_k}. \quad (\text{C.2})$$

The value of the gradient (C.1) depends on k , and when differentiating (7) with respect to ρ_k there are three different cases we must consider:

1. If $\rho_k = \rho_{12}$ the second term in (7) vanishes, because ρ_{12} is only present in \mathbf{R}_1 . Indeed,

$$\mathbf{R}^{11} = \begin{pmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{pmatrix},$$

which means that

$$\boldsymbol{\Sigma}^{(k)} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

and we easily see from (C.1) that in this particular case, $\{\nabla g(\boldsymbol{\rho})\}_k = 1$.

2. If $\rho_k = \rho_{ij}$ where $i \in \{1, 2\}$ and $j \in \{3, \dots, p\}$, then ρ_{ij} is an element in \mathbf{R}_2 , but not \mathbf{R}_1 or \mathbf{R}_3 . For that reason, we are left with the second term in (7), and furthermore, the middle term in the derivative (C.2) above will also vanish. The $2 \times (p-2)$ matrix $\partial \mathbf{R}_{12} / \partial \rho_{ij}$ will contain only zeros except for a 1 in the position of ρ_{ij} ; and $\partial \mathbf{R}^{21} / \partial \rho_{ij}$ is its transposed. Write $\mathbf{C} = (\mathbf{R}^{22})^{-1} \mathbf{R}^{21}$, and we see that $\Sigma^{(k)}$ in this case is a simple linear combination of the elements in \mathbf{C} :

$$\Sigma^{(k)} = \frac{\partial \mathbf{R}^{12}}{\partial \rho_{ij}} \mathbf{C} + \mathbf{C}^T \frac{\partial \mathbf{R}^{21}}{\partial \rho_{ij}},$$

from which we select the elements needed to calculate $\{\nabla g(\boldsymbol{\rho})\}_k$ through (C.1).

3. In this case, we look at the derivative of $g(\boldsymbol{\rho})$ with respect to the local correlations located in \mathbf{R}_3 , that is, ρ_{ij} for $i, j \in \{3, \dots, p\}$. We only need to keep the middle term in (C.2). We have that

$$\Sigma^{(k)} = -\mathbf{C}^T \frac{\partial \mathbf{R}_3}{\partial \rho_j} \mathbf{C},$$

where $\partial \mathbf{R}_3 / \partial \rho_k$ is a symmetric $(p-2) \times (p-2)$ matrix of zeros everywhere, except for 1s in the positions corresponding to ρ_k with respect to which we perform the differentiation. Finally, equation (C.2) provides the final expression for $\{\nabla g(\boldsymbol{\rho})\}_k$.

D PROOF OF THEOREM 3.2

Theorem 3.2 states that using marginally standard normal pseudo-observations instead of exactly standard normally distributed observations for estimating the LGPC does not change the conclusions in Theorem 3.1 in the main article and Theorem 2 in Otneim and Tjøstheim (2018). The result may be expected, as the empirical distribution functions \hat{F}_j , $j = 1, \dots, p$ converge at a rate faster than the local correlations ($n^{-\frac{1}{2}}$ vs. $(nb^5)^{-\frac{1}{2}}$ or $(nb^2)^{-\frac{1}{2}}$) depending on which theorem we refer to), but we must attend to some details nevertheless.

Proving joint asymptotic normality of the local Gaussian correlations with marginally normally distributed variables $(Z_1, \dots, Z_p) = (\Phi^{-1}(F_1(X_1)), \dots, \Phi^{-1}(F_p(X_p)))$, where F_j is the cdf of X_j for $i = 1, \dots, p$, relies on proving asymptotic normality for the variables

$$\mathbf{Y}_n(\mathbf{z}) = \frac{1}{n} \sum_{t=1}^n K_b(\mathbf{Z}_t - \mathbf{z}) \mathbf{v}(\mathbf{Z}_t, \boldsymbol{\rho}_0),$$

where $\mathbf{v}(\cdot) = \partial L(\cdot) / \partial \boldsymbol{\rho}$ in the same way as in Section B in this supplement. This has been done by Otneim and Tjøstheim (2017) in the iid case using arguments from Schervish (2012), and by Otneim and Tjøstheim (2018) in the α -mixing case, using arguments from Fan and Yao (2008). We need here to prove asymptotic normality of the variables

$$\widehat{\mathbf{Y}}_n(\mathbf{z}) = \frac{1}{n} \sum_{t=1}^n K_b(\widehat{\mathbf{Z}}_t - \mathbf{z}) \mathbf{v}(\widehat{\mathbf{Z}}_t, \boldsymbol{\rho}_0),$$

with $\widehat{\mathbf{Z}}_t = \Phi^{-1}(F_n(\mathbf{X}_t))$, and $F_n(\mathbf{X}_t)$ is the application of the marginal empirical distribution functions to the corresponding components of the vector \mathbf{X}_t .

In order to ease notation we state the proof in the simplified case only, where the local correlations are estimated using *pairs* (i, j) of variables. Then Y_n can be taken to be a scalar. It can be extended to the trivariate case by adding the appropriate terms in the derivations below, without changing the final conclusion. Write $\hat{Y}_n(\mathbf{z}) = Y_n(\mathbf{z}) - (Y_n(\mathbf{z}) - \hat{Y}_n(\mathbf{z}))$ and do a Taylor expansion of $Y_n(\mathbf{z})$ around $F_n(\mathbf{X}_t)$. Since we use product kernels, we then need derivatives of (assuming without loss of generality that $K(x) = K(-x)$)

$$H_{\mathbf{z},b} = K\left(\frac{z_1 - \Phi^{-1}(y_1)}{b}\right) K\left(\frac{z_2 - \Phi^{-1}(y_2)}{b}\right) v(\Phi^{-1}(\mathbf{y})),$$

where $\mathbf{y} = F(\mathbf{x})$, $y_i = F(x_i)$, $i = 1, 2$, and we write $K(x_1, x_2) = K(x_1)K(x_2)$. Writing $k(z) = K'(z)$ and $v_i = \partial v_i / \partial y_i$,

$$\begin{aligned} \frac{\partial H_{\mathbf{z},b}}{\partial y_1} &= k\left(\frac{z_1 - \Phi^{-1}(y_1)}{b}\right) K\left(\frac{z_2 - \Phi^{-1}(y_2)}{b}\right) \frac{v(\Phi^{-1}(\mathbf{y}))}{b\phi(\Phi^{-1}(y_1))} \\ &\quad + K\left(\frac{z_1 - \Phi^{-1}(y_1)}{b}\right) K\left(\frac{z_2 - \Phi^{-1}(y_2)}{b}\right) \frac{v_1(\Phi^{-1}(\mathbf{y}))}{\phi(\Phi^{-1}(y_1))}, \\ \frac{\partial H_{\mathbf{z},b}}{\partial y_2} &= K\left(\frac{z_1 - \Phi^{-1}(y_1)}{b}\right) k\left(\frac{z_2 - \Phi^{-1}(y_2)}{b}\right) \frac{v(\Phi^{-1}(\mathbf{y}))}{b\phi(\Phi^{-1}(y_2))} \\ &\quad + K\left(\frac{z_1 - \Phi^{-1}(y_1)}{b}\right) K\left(\frac{z_2 - \Phi^{-1}(y_2)}{b}\right) \frac{v_2(\Phi^{-1}(\mathbf{y}))}{\phi(\Phi^{-1}(y_2))}. \end{aligned}$$

A typical term in the Taylor expansion of $Y_n(\mathbf{z})$ takes the following form:

$$\begin{aligned} K_b\left(\Phi^{-1}(F(\mathbf{X}_t)) - \mathbf{z}\right) v\left(\Phi^{-1}(F(\mathbf{X}_t)), \rho_0\right) &= \\ K_b\left(\Phi^{-1}(F_n(\mathbf{X}_t)) - \mathbf{z}\right) v\left(\Phi^{-1}(F_n(\mathbf{X}_t)), \rho_0\right) &+ \\ + k\left(\frac{z_1 - \Phi^{-1}(F_n^*(X_{1,t}))}{b}\right) K\left(\frac{z_2 - \Phi^{-1}(F_n^*(X_{2,t}))}{b}\right) \frac{v(\Phi^{-1}(F_n^*(\mathbf{X}_t)))}{b\phi(\Phi^{-1}(F_n^*(X_{1,t})))} (F(X_{1,t}) - F_n(X_{1,t})) &+ \\ + K\left(\frac{z_1 - \Phi^{-1}(F_n^*(X_{1,t}))}{b}\right) K\left(\frac{z_2 - \Phi^{-1}(F_n^*(X_{2,t}))}{b}\right) \frac{v_1(\Phi^{-1}(F_n^*(\mathbf{X}_t)))}{\phi(\Phi^{-1}(F_n^*(X_{1,t})))} (F(X_{1,t}) - F_n(X_{1,t})) &+ \\ + \text{analogous terms for the second (and possibly third) index involving } (F(X_{2,t}) - F_n(X_{2,t})), \end{aligned}$$

where F_n^* comes from the mean value theorem. The challenge in proving the desired result is to control the behavior of quantities on the form $1/\phi(\Phi^{-1}(F(x)))$, because this fraction tends to infinity as $|x| \rightarrow \infty$. The key to this problem is the assumption that the kernel function $K(\cdot)$, and thus its derivative k , have compact support. This means, that in the expressions above, $K = k \equiv 0$ if for some $M > 0$

$$\frac{|z_1 - \Phi^{-1}(F_n^*(X_{1,t}))|}{b} \geq M,$$

or

$$|z_1 - \Phi^{-1}(F_n^*(X_{1,t}))| \geq Mb,$$

or, by removing the absolute value signs,

$$\Phi^{-1}(F_n^*(X_{1,t})) \leq z_1 - Mb \quad \text{or} \quad \Phi^{-1}(F_n^*(X_{1,t})) \geq z_1 + Mb,$$

where M may be large. The same reasoning applies of course to the second index as well. Letting $n \rightarrow \infty$ and using consistency of the empirical distribution function, it follows that the kernel function term is zero if

$$X_{1,t} \leq F^{-1}(\Phi(z_1 - Mb)) \quad \text{or} \quad X_{1,t} \geq F^{-1}(\Phi(z_1 + Mb)).$$

The same reasoning applies to the derivative of the kernel function, where we consider the function k/b instead of K . All of this implies that the challenge of controlling the magnitude of the Taylor terms above as $x \rightarrow \infty$ disappears, since, taking boundedness of the other terms into account,

$$k\left(\frac{z_1 - \Phi^{-1}(F_n^*(X_{t,1}))}{b}\right) K\left(\frac{z_2 - \Phi^{-1}(F_n^*(X_{t,2}))}{b}\right) \frac{v(\Phi^{-1}(F_n^*(\mathbf{X}_t)))}{b\phi(\Phi^{-1}(F_n^*(X_{t,1})))}$$

and

$$K\left(\frac{z_1 - \Phi^{-1}(F_n^*(X_{t,1}))}{b}\right) K\left(\frac{z_2 - \Phi^{-1}(F_n^*(X_{t,2}))}{b}\right) \frac{v_1(\Phi^{-1}(F_n^*(\mathbf{X}_t)))}{\phi(\Phi^{-1}(F_n^*(X_{t,1})))}$$

are bounded almost surely as $n \rightarrow \infty$ and $b \rightarrow 0$.

All this means that if we can prove that for $i = 1, 2$,

$$\frac{1}{n} \sum_{t=1}^n (F_n(X_{i,t}) - F(X_{i,t})) \xrightarrow{\mathbb{P}} 0$$

at an appropriate rate, then we are done by Slutsky's Theorem.

In the original version of this proof we used an argument based on the law of the iterated logarithm of Rio (1995), leading to a rate of $n^{-1/2}(2 \log \log n)^{1/2}$, but an anonymous referee provided us with a much simpler argument based on a U-statistic decomposition for a strongly mixing process:

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n (F_n(X_{i,t}) - F(X_{i,t})) &= \frac{1}{n^2} \sum_{t=1}^n \sum_{s \neq t}^n (I(X_{i,s} \leq X_{i,t}) - F(X_{i,t})) + \frac{1}{n^2} \sum_{t=1}^n (1 - F(X_{i,t})) \\ &= \frac{1}{n} \sum_{t=1}^n (0.5 - F(X_{i,t})) + O_p(n^{-1}) \\ &= O_p(n^{-1/2}) + O_p(n^{-1}) \\ &= O_p(n^{-1/2}). \end{aligned}$$

E THE VALIDITY OF THE BOOTSTRAP FOR ESTIMATION

The asymptotic theory for the LGPC involves complicated expressions that may be difficult to implement in practice. It is then natural to use the bootstrap instead to construct confidence

intervals. In order to accommodate the time series case, we must use the block bootstrap, and we are thus faced with the problem of proving the validity of the block bootstrap.

We have seen that the building blocks of the asymptotic estimation theory of the LGPC between the vectors \mathbf{Z}_1 and \mathbf{Z}_2 given \mathbf{Z}_3 are the pairwise local Gaussian correlations contained in the local correlation matrix Σ . Let us simply use the notation $\rho_{ij} = \rho_{ij}(z_i, z_j)$ for one of these local correlations and $\hat{\rho}_{ij}$ for its estimate as outlined in the main article. If we can prove that the block bootstrap is valid for such a pairwise local likelihood estimate, then it follows fairly straightforwardly that it is also valid for the appropriate vector of such pairwise estimates used in the asymptotic theory laid out above. Note that the 3-coordinate case (i) in Section A.3 requires essentially the same arguments.

We can use the bootstrap theory by Lacal and Tjøstheim (2017, 2019) more or less directly, which in turn is based on Gonçalves and White (2002, 2004), but we spell out some details below. First, we need to specify the analogue of the conditions A1–A5 by Lacal and Tjøstheim (2019), where we adapt their notation to our situation by replacing the parameter θ with ρ_{ij} , and the time series $\{X_t, Y_t\}$ by our multivariate observations $\{\mathbf{Z}_t\}$. Since $\rho = \rho_{ij}$ is estimated using the two corresponding data vectors only, we can, for the time being and without loss of generality, assume that the observation series is bivariate: $\mathbf{Z}_t = \{Z_{i,t}, Z_{j,t}\}$. Note further that it suffices to consider the true pseudo-observations $\{\mathbf{Z}_t\}$ instead of the empirical pseudo observations $\{\widehat{\mathbf{Z}}_t\}$ in the following because the difference between them is asymptotically negligible according to Theorem 3.2 in the main article.

To state the bootstrap results we need a few more concepts and three additional conditions. We block-bootstrap the entire vector series $\{\mathbf{Z}_t\}$. This automatically leads to the same block-bootstrap for any pair of components which we have denoted by $\{Z_{i,t}, Z_{j,t}\}$ above. Lacal and Tjøstheim (2019) use both the block bootstrap as well as the stationary bootstrap, and the results below hold irrespective of this particular choice. We will therefore not distinguish between them here, but rather refer the reader to Lacal and Tjøstheim (2019) for further details on this distinction. Bootstrapped quantities will be denoted with a star (*) in the following.

Before we introduce the specific bootstrap conditions **A4** and **A5** below, we need to state the following condition. The notation of equations (A.13)–(A.16) is used.

A3 For a fixed bandwidth b , and using stationarity,

$$J_{n,b} = \mathbb{E} \left(-\nabla^2 L_{n,b}(\rho_{0,b}) \right) = J_b = \mathbb{E} \left(-\nabla^2 L^{(b)}(Z_{i,t}, Z_{j,t}; \rho_{0,b}) \right)$$

is of order $O(1)$, and continuous on Θ uniformly in n . Further,

$$M_{n,b} = \text{Var} \left(\sqrt{nb_1 b_2} \nabla L_{n,b}(\rho_0, b) \right)$$

is $O(1)$ and uniformly positive. Here ∇ is the gradient with respect to ρ , which in the present scalar parameter case is simply the differential operator $d/d\rho$.

The bootstrap lives on the probability space $(\Lambda, \mathcal{G}, \mathbb{P}^*)$. The outcomes $\lambda \in \Lambda$ depend on the outcomes $\omega \in \Omega$ in the original probability space, and we can now state two more assumptions:

- A4** The bootstrap probability space $(\Lambda, \mathcal{G}, \mathbb{P}^*)$ is a complete probability space for all $\omega \in \Omega$. Let $L_{n,b}$ be defined as in (A.13) with $\rho \in \Theta$. Let $\{L_{n,b}^* : \Lambda \times \Omega \times \Theta \rightarrow \overline{\mathbb{R}}\}$ be a sequence of random functions such that $L_{n,b}^*(\rho) = L_{n,b}((\mathbf{Z}_{in}^*(\lambda, \omega), \mathbf{Z}_{jn}^*(\lambda, \omega)), \rho)$ where $Z_{i,t}^*(\lambda, \omega) = Z_{i\tau_t^Z(\lambda)}(\omega)$, $Z_{j,t}^*(\lambda, \omega) = Z_{j\tau_t^Z(\lambda)}(\omega)$, where $\tau_t^Z : \Lambda \rightarrow \mathbb{N}$, $\omega \in \Omega$, $\lambda \in \Lambda$ are vectors of random indices representing the block bootstrap operation on $\{\mathbf{Z}_t\}$. Moreover, the block length l_n is such that $l_n = o(\sqrt{n})$ as $n \rightarrow \infty$.
- A5** For $\rho \in \Theta$ and for every $t = 1, \dots, n$, $\{L^{(b)}(Z_{i,t}, Z_{j,t}), \rho\}$ is Lipschitz continuous on Θ , $\{\nabla L^{(b)}((Z_{i,t}, Z_{j,t}), \rho)\}$ is 6-dominated on Θ uniformly in t, n , and $\{\nabla^2 L^{(b)}((Z_{i,t}, Z_{j,t}), \rho)\}$ is Lipschitz continuous on Θ and 2-dominated on Θ uniformly in t, n (r -dominated is defined after (A.16)).

Assumption A4 sets the stage for the block bootstrapping, whereas A5 is needed in proving the validity of the bootstrap central limit theorem. The assumption A4 is not restrictive, and again we refer to Lacal and Tjøstheim (2019) for more discussion.

In order to state the bootstrap validity theorems, we need to specify convergence in the space $(\Lambda, \mathcal{G}, \mathbb{P}_\omega^*)$. This is done as by Gonçalves and White (2002, 2004). First, for a random variable Y_n^* , we write $Y_n^* \xrightarrow{\mathbb{P}_\omega^*, \mathbb{P}} 0$ if for any $\varepsilon, \delta > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(\omega : \mathbb{P}_\omega^*(\lambda : |Y_n^*(\lambda, \omega)| > \varepsilon) > \delta) = 0.$$

Further, we write $Y_n^* \xrightarrow{d_{\mathbb{P}_\omega^*}} \mathcal{N}(0, 1)$ prob $-\mathbb{P}$, if for every sub-sequence $\{n'\}$, there exists a further sub-sequence $\{n''\}$ such that $Y_{n''}^* \xrightarrow{d_{\mathbb{P}_\omega^*}} \mathcal{N}(0, 1)$ almost surely, see Gonçalves and White (2004, page 210). This definition is based on the fact that convergence in probability implies almost sure convergence for such kinds of sub-sequences (see Theorem 20.5 of Billingsley (2008)).

We proceed by stating the validity of the bootstrap in terms of consistency and a central limit theorem:

Theorem E.1. *Let assumptions A1, A2 and A4 hold, and let $\{L^{(b)}((Z_{i,t}, Z_{j,t}), \rho)\}$, $\rho \in \Theta$ be Lipschitz continuous and 4-dominated on Θ uniformly in t, n . Moreover, assume that, as $b = b_n \rightarrow 0$, $\sigma_b^2 \stackrel{\text{def}}{=} \text{Var}(L^{(b)}((Z_{i,t}, Z_{j,t}), \rho)) \rightarrow \sigma^2 < \infty$ and $\mu_b \stackrel{\text{def}}{=} E(L^{(b)}(Z_{i,t}, Z_{j,t}), \rho) \rightarrow \mu < \infty$, for all $t = 1, \dots, n$ and $\rho \in \Theta$. Then,*

$$\rho_{n,b}^* - \rho_{b,n} \xrightarrow{\mathbb{P}_\omega^*, \mathbb{P}} 0. \quad (\text{E.1})$$

Theorem E.2. *Under the conditions 1–5 of Theorems 3.1 and 3.2 in the main article and assumptions A1–A5 stated above,*

$$\sqrt{nb_1 b_2 M_{n,b}^{-1/2}} J_{n,b}(\rho_{n,b}^* - \rho_{n,b}) \xrightarrow{d_{\mathbb{P}_\omega^*}} \mathcal{N}(0, 1) \text{ prob} - \mathbb{P}. \quad (\text{E.2})$$

Theorems E.1 and E.2 are formulated in their simplest case here, which concerns only one local parameter. A more complicated version is given by Lacal and Tjøstheim (2019) concerning the five parameter case. The proofs are essentially the same and are spelled out in some detail in the supplementary material of that paper as the proofs of Theorems D1 and D2.

Theorems E.1 and E.2 concern only one local correlation. The local *partial* correlation, however, is a function of a vector of local correlations. The form of this function and the vector of local correlations depend on the generality and the type of model considered as seen from the alternatives (i) – (v) in Section A.3 of this supplement. Corollaries E.3 and E.4 cover the general situation with pairwise model and pairwise coordinates, including alternative (i)', which is alternative (i) but restricted to pairwise coordinates. Corollary E.3 treats bootstrap consistency and Corollary E.4 gives a bootstrap CLT:

Corollary E.3. *Assume the conditions of Theorem E.1. Let $\alpha_{n,b}(i, j|\mathbf{Z}_3) \stackrel{\text{def}}{=} \hat{\rho}(Z_{1i}, Z_{2j}|\mathbf{Z}_3)$ be the i, j th element of the estimated partial correlation matrix of $(\mathbf{Z}_1, \mathbf{Z}_2)$ given \mathbf{Z}_3 . This depends on scalar coordinates (z_1, z_2) and a vector coordinate \mathbf{z}_3 . Note that $\alpha_{n,b}(i, j|\mathbf{Z}_3)$ is a continuous function of a vector of local correlations according to any of the alternatives (i)', (ii)–(v) in Section A.3. Then $\alpha_{n,b}(i, j|\mathbf{Z}_3) - \alpha_{n,b}^*(i, j|\mathbf{Z}_3) \xrightarrow{\mathbb{P}_{\omega}^*, \mathbb{P}} 0$.*

Proof. Denote by $\boldsymbol{\rho} = \boldsymbol{\rho}(\mathbf{Z})$ the vector of local correlations according to one of the alternatives (i)', (ii)–(v). This vector has a finite number of components and equation (E.1) guarantees the bootstrap consistency of each component. The bootstrap consistency for the vector follows immediately and the proof is concluded by the continuous mapping theorem. ■

The next results gives the asymptotic normality of the bootstrapped local partial correlation. We let $\alpha(i, j|\mathbf{Z}_3) = g_{ij}(\boldsymbol{\rho}(\mathbf{Z}))$ be the local partial correlation as defined above and expressed as a function g_{ij} of the vector of local correlations according to one of the alternatives (i)', (ii)–(v). Using the notation of Section A.3 of this supplement, the correlation vector $\boldsymbol{\rho}(\mathbf{Z})$ will in general have elements consisting of $\rho_{ij}^{1,2}, i = 1, \dots, d_1, j = 1, \dots, d_2; \rho_{ij}^{1,3}, i = 1, \dots, d_1, j = 1, \dots, d_3; \rho_{ij}^{2,3}, i = 1, \dots, d_2, j = 1, \dots, d_3$, where d_1, d_2, d_3 are chosen according to one of the alternatives (i)', (ii)–(v).

Corollary E.4. *Let the assumptions of Theorem E.2 hold and let $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}(\boldsymbol{\rho})$ be the covariance matrix of the $\hat{\boldsymbol{\rho}}$ vector. Then,*

$$\sqrt{nb_1b_2}(\alpha_{n,b}^*(i, j) - \alpha_{n,b}(i, j)) \xrightarrow{d_{\mathbb{P}_{\omega}^*}} \mathcal{N}(0, \nabla g(\boldsymbol{\rho})^T \boldsymbol{\Lambda}(\boldsymbol{\rho}) \nabla g(\boldsymbol{\rho})) \text{ prob} - \mathbb{P}.$$

Proof: The proof of Theorem E.2 is given as the proof of Theorem D2 in the online appendix accompanying Lacal and Tjøstheim (2019). This proof can be adapted to prove Corollary E.4. The starting point consists in replacing $L_{n,b}^*$ and $L_{n,b}$ by $L^* = \sum_{i,j} a_{ij} L_{ij,n,b}^*$ and $L = \sum_{i,j} a_{ij} L_{ij,n,b}$, respectively, where the sum runs over all possible pairs (i, j) corresponding to $\rho_{ij}^{1,2}, \rho_{ij}^{1,3}, \rho_{ij}^{2,3}$ that are included in the $\boldsymbol{\rho}$ -vector. The derivations in the proof of Theorem D2 by Lacal and Tjøstheim (2019) deal with the more complicated five-parameter case. In the present case, however, the vector $\boldsymbol{\rho}$ consists of correlations ρ_{ij} that have been estimated using

the one-parameter pairwise setup, where, cf. Otneim and Tjøstheim (2017), the quantities J_{ij} and M_{ij} are scalars. Here $L_{ij,n,b}$ is the log likelihood for the pair (i, j) treated in the proof of Theorem D2, and a_{ij} is a corresponding sequence of real numbers. Because of the linearity of the expressions for L^* and L it is easily checked by going step by step that the proof of Theorem D2 of Lacal and Tjøstheim (2019) can be carried through for L^* and L . The only difference is that the quantities $J_{n,b}$ and $M_{n,b}$ now have to be replaced by $J_{n,b}(a) = E(-\sum_{i,j} a_{ij} \nabla^2 L_{ij,n,b}(\rho_{0,b}))$ and $\text{Var}(\sqrt{nb_1 b_2} \sum_{i,j} a_{ij} \nabla L_{ij,n,b})$. This results in joint asymptotic bootstrap normality using the Cramer-Wold device. Applying the delta rule now proves the Corollary. ■

Let \mathbf{M}_{ij} and \mathbf{J}_{ij} be as in the proof of Corollary E.4 for a fixed ρ_{ij} . Note that in the one-parameter case of ρ_{ij} in Corollary E.4, to first order the covariance matrix $\mathbf{\Lambda}$ is a diagonal matrix having ordered values of M_{ij}/J_{ij}^2 along the diagonal. This can be seen using the methods of proofs of Theorem 2 of Otneim and Tjøstheim (2017) and Theorem 2 Otneim and Tjøstheim (2018). Note also that the 3-coordinate case of alternative (i) can be stated and proved in the same way, but a different scaling factor is required to avoid a singular covariance matrix as seen in the proof of Theorem 3.1 in the main article. The 3-coordinate case is in fact closer to the situation in Lacal and Tjøstheim (2019).

F ASYMPTOTIC THEORY FOR THE TEST STATISTIC

Consider the general case $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3)$ where each component is d_i -dimensional correspondingly, with $i = 1, 2, 3$, and \mathbf{Z} , naturally, is $d_1 + d_2 + d_3$ -dimensional, and consider also the local partial correlation matrix of $(\mathbf{Z}_1, \mathbf{Z}_2 | \mathbf{Z}_3)$. In alternative (iii), where $d_3 = 1$, in Section A.3 this matrix is given by equation (A.10) with

$$\alpha_{ij} = \rho(Z_{1,i}, Z_{2,j} | Z_3), \quad i = 1, \dots, d_1; j = 1, \dots, d_2.$$

This forms a $d_1 \times d_2$ matrix $\boldsymbol{\alpha}$. One can test for conditional independence between $Z_{1,i}$ and $Z_{2,j}$ given \mathbf{Z}_3 by testing whether $\alpha_{ij} = 0$, or one can test for conditional independence between \mathbf{Z}_1 and \mathbf{Z}_2 given \mathbf{Z}_3 by testing whether $\boldsymbol{\alpha} = \{\alpha_{ij}\} = \mathbf{0}$. The latter case is in a sense the most general, and we will stick to that in the sequel. One can then test for $\boldsymbol{\alpha} = \mathbf{0}$ by using a matrix norm, for instance the Frobenius norm, where $\|\boldsymbol{\alpha}\| = \sqrt{\sum_{i,j} \alpha_{ij}^2} = 0$. A more convenient test statistic is $\sum_{i,j} \alpha_{ij}^2$. To follow the notation of Lacal and Tjøstheim (2019) we write, letting $\mathbf{z} = (\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3)$,

$$T = \int_S h(\boldsymbol{\rho}(\mathbf{z})) dF(\mathbf{z})$$

where S is an area of integration, F is the cumulative distribution function of \mathbf{Z} , and $\boldsymbol{\rho} = \boldsymbol{\rho}(\mathbf{z})$, as before, is a vector consisting of all possible pairs that can be formed by picking pairwise components from \mathbf{Z} . The number of components of this vector depends on which of the alternatives models (i)–(v) in Section A.3 one considers. It is reasonable to express T as a function of $\boldsymbol{\rho}$ since all partial correlations can be expressed as a function of this vector (see e.g. again equation (A.10)). One possible choice of the function h is $h(\boldsymbol{\rho}(\mathbf{z})) = \sum_{i,j} \alpha_{i,j}^2(z_{1,i}, z_{2,j} | \mathbf{z}^{(2)})$, which is a function of $\boldsymbol{\rho}$ since the α -s are functions of $\boldsymbol{\rho}$.

The corresponding function in Lacal and Tjøstheim (2019) is $h(\theta(\mathbf{x}), \mathbf{x})$, where $\theta(\mathbf{x})$ is at most 5-dimensional and \mathbf{x} is bivariate. As in that paper we proceed in two stages. First we let the bandwidth vector \mathbf{b} be fixed and let $n \rightarrow \infty$, and then we let $\mathbf{b} = (b_1, b_2) = (b, b) \rightarrow \mathbf{0}$. Accordingly we define

$$T_b = \int_S h(\boldsymbol{\rho}_b(\mathbf{z})) dF(\mathbf{z})$$

where $\boldsymbol{\rho}_b$ is the population quantity for a fixed bandwidth \mathbf{b} , and $T_{n,b}$ by

$$T_{n,b} = \int_S h(\boldsymbol{\rho}_{b,n}(\mathbf{z})) dF_n(\mathbf{z}),$$

where $\boldsymbol{\rho}_{b,n}(\mathbf{z})$ is the estimate of $\boldsymbol{\rho}(\mathbf{z})$ obtained by pairwise likelihood for each component of $\boldsymbol{\rho}(\mathbf{z})$.

Consistency is straightforward to prove:

Theorem F.1. *Let the conditions of Theorem A.1 be fulfilled. Then $T_{n,b} \xrightarrow{\mathbb{P}} T_b$ as $n \rightarrow \infty$ and $T_{n,b} \xrightarrow{\mathbb{P}} T$ as $n \rightarrow \infty$ and $\mathbf{b} \rightarrow \mathbf{0}$.*

Proof. This result follows directly from Theorem A.1 and the continuous mapping theorem. ■

It is more difficult to obtain the limiting distribution. We will follow the technique in the proof of Theorem 3.2 in Lacal and Tjøstheim (2019). See also Joe (1989). The details are a bit more involved in the present case because of the larger dimension and because we need to involve the pairwise log likelihood in the procedure.

The results on asymptotic distribution can be proved on several levels. We consider two cases: a) the pairwise simplification where involving the estimates $\rho_{ij,n,b}(z_i, z_j)$ for all possible pairs (i, j) of the $\boldsymbol{\rho}$ vector, and b) the case where $d_1 = d_2 = d_3 = 1$, but where $\rho_{ij} = \rho_{ij}(\mathbf{z})$ is allowed to depend on all three coordinates $\mathbf{z} = (z_1, z_2, z_3)$. We will put the emphasis on a) and only comment briefly on b).

Theorem F.2. *Let the assumptions A, D–F of Theorem 3 of Otneim and Tjøstheim (2018) and the conditions of Theorem 3.2 of Lacal and Tjøstheim (2018) (which includes conditions A1 and A2 of this supplement) be fulfilled. Then, as $n \rightarrow \infty$,*

$$\sqrt{n} \left(C_n(A_b) \right)^{-\frac{1}{2}} (T_{n,b} - T_b) \xrightarrow{d} \mathcal{N}(0, 1),$$

where

$$A_b(\mathbf{z}) = \int_S \left[\sum_{i,j} \left\{ \frac{\partial}{\partial \rho_{ij}} h(\boldsymbol{\rho}_b, (\mathbf{v})) \right\} J_{ij}^{-1}(\rho_{ij,b}(\mathbf{v}_{ij})) K_b(\mathbf{z}_{ij} - \mathbf{v}_{ij}) u_{ij}(\mathbf{z}_{ij}, \rho_{ij,b}(\mathbf{v}_{ij})) \right] dF(\mathbf{v}) \\ + h(\boldsymbol{\rho}_b(\mathbf{z})) 1_S(\mathbf{z})$$

and

$$C_n(A_b) = \int A_b^2(\mathbf{z}) dF(\mathbf{z}) + \frac{1}{n} \sum_{s \neq t} \int A_b(\mathbf{z}) A_b(\mathbf{u}) dF^{s,t}(\mathbf{z}, \mathbf{u}) \\ - n \int A_b(\mathbf{z}) A_b(\mathbf{u}) dF(\mathbf{z}) dF(\mathbf{u}) \quad (\text{F.1})$$

with J_{ij} being the bivariate version of \mathbf{J}_b defined in Theorem 3.1, and, with $F^{s,t}(\mathbf{z}, \mathbf{u})$ being the joint cumulative distribution function of $(\mathbf{Z}_s, \mathbf{Z}_t)$, and where $\{\mathbf{Z}_t, t = 1, \dots, n\}$ are the n observations of \mathbf{Z} and $u_{ij}(\mathbf{Z}_{ij}, \rho_{ij,b}(\mathbf{v}_{ij})) = \frac{\partial}{\partial \rho_{ij,b}} \log \psi(\mathbf{Z}_{ij}, \rho_{ij,b}(\mathbf{v}_{ij}))$.

To ease comparison with Lacal and Tjøstheim (2018) we have written and will write $\mathbf{z}_{ij} \stackrel{\text{def}}{=} (z_i, z_j)$ and $\mathbf{v}_{ij} \stackrel{\text{def}}{=} (v_i, v_j)$. Further, if $n \rightarrow \infty$ and $\mathbf{b}_n \rightarrow \mathbf{0}$, the relationship between them being as in Assumption 3 of Theorem 3.1 in the main article, then T_b can be replaced by T ,

$$\sqrt{n}(C_n(A_b))^{-\frac{1}{2}}(T_{n,b} - T) \xrightarrow{d} \mathcal{N}(0, 1).$$

Note that the sum in the second term in the definition of $C_n(\mathbf{A}_b)$ taken together with the third term converges due to the geometric α -mixing condition of A1. It is also important to note that conditional independence is not assumed in Theorem F.2. This will be exploited in Section G. See also Theorem 3.2 in Lacal and Tjøstheim (2019) and its relation to Corollary 3.1 in that paper.

Proof. From the proof of Theorem 3.1 in Lacal and Tjøstheim (2017), one has that asymptotically,

$$T_{n,b} - T_b \sim \frac{1}{\sqrt{n}} \left\{ \int_S h(\boldsymbol{\rho}_b(\mathbf{z})) dG_n(\mathbf{z}) + \int_S \nabla_{\boldsymbol{\rho}} h(\boldsymbol{\rho}_b(\mathbf{z}))^T \mathbf{V}_n(\mathbf{z}) dF(\mathbf{z}) \right\}. \quad (\text{F.2})$$

Here

$$G_n(\mathbf{z}) = \sqrt{n}(F_n(\mathbf{z}) - F(\mathbf{z})) \quad \text{and} \quad \mathbf{V}_n(\mathbf{z}) = \sqrt{n}((\boldsymbol{\rho}_{n,b}(\mathbf{z}) - \boldsymbol{\rho}_b(\mathbf{z})),$$

with F_n being the empirical cumulative function of \mathbf{Z} , and where, again, $\boldsymbol{\rho}$ is the vector consisting of all possible pairs ρ_{ij} . This rewriting of $T_{n,b} - T_b$ is the first step in obtaining an expression in terms over \mathbf{Z}_t , $t = 1, \dots, n$ so that a mixing CLT can be used. We proceed to the pairwise analysis.

Let (i, j) be a fixed pair and $\rho_{ij,n,b}(\mathbf{z}_{ij})$ be the corresponding estimated local correlation for the pair (i, j) chosen from the vector $\boldsymbol{\rho}_{n,b}$. Here the pair of coordinates $\mathbf{z}_{ij} = (z_i, z_j)$ corresponds to the bivariate \mathbf{x} in the proof of Theorem 3.1 by Lacal and Tjøstheim (2017). We will also omit some indices or parts of indices when it is clear what is meant. Finally, we omit the area of integration S . We are using the pairwise local log likelihood given by

$$L_{ij,n} = \frac{1}{n} \sum_{t=1}^n K_b(\mathbf{Z}_{ij,t} - \mathbf{z}_{ij}) \log \psi(\mathbf{Z}_{ij,t}, \rho_{b,ij}(\mathbf{z}_{ij})) - \int K_b(\mathbf{v}_{ij} - \mathbf{z}_{ij}) \psi_{ij}(\mathbf{v}_{ij}, \rho_{b,ij}(\mathbf{z}_{ij})) d\mathbf{v}_{ij},$$

where $K_b = b_1^{-1} K_1(\cdot/b_1) b_2^{-1} K_2(\cdot/b_2)$ is a product kernel function, and $\mathbf{Z}_{ij,t} = (Z_{i,t}, Z_{j,t})$ is the t -th pair of observations, and $\mathbf{v}_{ij} = (v_i, v_j)$. Estimates $\rho_{ij,b,n}(\mathbf{z}_{ij})$ are obtained by differentiation of the log likelihood and by introducing the score function U_{ij} and information function I_{ij} as follows

$$\begin{aligned} U_{ij}(\mathbf{z}_{ij}, \rho_{ij}) &= \frac{\partial}{\partial \rho_{ij}} L_{ij,n}(\mathbf{z}_{ij}, \rho_{ij}) \\ &= \frac{1}{n} \sum_{t=1}^n K_b(\mathbf{Z}_{ij,t} - \mathbf{z}_{ij}) u_{ij}(\mathbf{Z}_{ij,t}, \rho_{ij}) - \int K_b(\mathbf{v}_{ij} - \mathbf{z}_{ij}) u_{ij}(\mathbf{v}_{ij}, \rho_{ij}) \psi(\mathbf{v}_{ij}, \rho_{ij}) d\mathbf{v}_{ij} \end{aligned}$$

where $u_{ij}(\mathbf{z}_{ij}, \rho_{ij}) = \partial/\partial\rho_{ij} \log \psi(\mathbf{z}_{ij}, \rho_{ij})$. Similarly

$$I_{ij}(\mathbf{z}_{ij}, \rho_{ij}) = \frac{\partial^2}{\partial\rho_{ij}^2} L_{ij,n}(\mathbf{z}_{ij}, \rho_{ij}) \quad \text{and} \quad J_{ij,b}(\mathbf{z}_{ij}) = \mathbb{E}(-I_{ij}(\mathbf{z}_{ij}, \rho_{ij})),$$

where Lacal and Tjøstheim (2017, 2019) may be consulted for more details in the derivation and a more explicit expression for $I_{ij}(\mathbf{z}_{ij}, \rho_{ij})$ for the case considered there. In contradistinction to the 5-parameter case considered by Lacal and Tjøstheim (2019), in our present 1-parameter case $J_{b,ij}$ is not singular as $\mathbf{b} \rightarrow \mathbf{0}$. In fact, it is shown by Otneim and Tjøstheim (2017, Theorem 2) that as $\mathbf{b} \rightarrow \mathbf{0}$, $J_{b,ij} \rightarrow u(\mathbf{z}_{ij}, \rho_{ij})\psi(\mathbf{z}_{ij}, \rho(\mathbf{z}_{ij}))$.

The vector $\mathbf{V}_n(\mathbf{z}) = \sqrt{n}[\boldsymbol{\rho}_{n,b}(\mathbf{z}) - \boldsymbol{\rho}_b(\mathbf{z})]$ has the dimension of $\boldsymbol{\rho}(\mathbf{z})$ and consists of components $\{\rho_{ij,n,b}(\mathbf{z}_{ij}) - \rho_{ij,b}(\mathbf{z}_{ij})\}$ where i and j in the most general case run over all possible pairs formed by the components of $\boldsymbol{\rho}(\mathbf{z})$. To be able to use the mixing central limit theorem we follow Lacal and Tjøstheim (2019) and express this more explicitly in terms of the variables \mathbf{Z}_t , $t = 1, \dots, n$. By Taylor expansion as in Lacal and Tjøstheim (2019),

$$\begin{aligned} 0 &= U_{ij}(\mathbf{z}_{ij}, \rho_{ij,n,b}(\mathbf{z}_{ij})) \sim U_{ij}(\mathbf{z}_{ij}, \boldsymbol{\rho}_{ij,b}(\mathbf{z}_{ij})) + I_{ij}(\mathbf{z}_{ij}, \rho_{ij,b}(\mathbf{z}_{ij}))(\rho_{ij,n,b}(\mathbf{z}_{ij}) - \rho_{ij,b}(\mathbf{z}_{ij})) \\ &\sim U_{ij}(\mathbf{z}_{ij}, \rho_{ij,b}(\mathbf{z}_{ij})) - J_{ij,b}(\mathbf{z}_{ij})(\rho_{ij,n,b}(\mathbf{z}_{ij}) - \rho_{ij,b}(\mathbf{z}_{ij})), \end{aligned}$$

which implies

$$\begin{aligned} \rho_{ij,n,b}(\mathbf{z}_{ij}) - \rho_{ij,b}(\mathbf{z}_{ij}) &\sim J_{ij,b}^{-1}(\mathbf{z}_{ij})U_{ij}(\mathbf{z}_{ij}, \rho_{ij,b}(\mathbf{z}_{ij})) \\ &= J_{ij,b}^{-1}(\mathbf{z}_{ij}) \left[\frac{1}{n} \sum_{t=1}^n K_b(\mathbf{Z}_{ij,t} - \mathbf{z}_{ij}) u_{ij}(\mathbf{Z}_{ij,t}, \rho_{ij,b}(\mathbf{z}_{ij})) \right. \\ &\quad \left. - \int K_b(\mathbf{v}_{ij} - \mathbf{z}_{ij}) u_{ij}(\mathbf{v}_{ij}, \rho_{ij,b}(\mathbf{z}_{ij})) \psi(\mathbf{v}_{ij}, \rho_{ij,b}(\mathbf{z}_{ij})) d\mathbf{v}_{ij} \right]. \end{aligned} \quad (\text{F.3})$$

The second term in (F.2) can then be expressed as

$$\begin{aligned} &\int_S \sum_{i,j} \frac{\partial}{\partial\rho_{ij}} h(\rho_{ij,b}(\mathbf{z})) \\ &\quad \times J_{ij,b}^{-1}(\mathbf{z}_{ij}) \left[\frac{1}{n} \sum_{t=1}^n K_b(\mathbf{Z}_{ij,t} - \mathbf{z}_{ij}) u_{ij}(\mathbf{Z}_{ij,t}, \rho_{ij,b}(\mathbf{z}_{ij})) \right. \\ &\quad \left. - \int K_b(\mathbf{v}_{ij} - \mathbf{z}_{ij}) u_{ij}(\mathbf{v}_{ij}, \rho_{ij,b}(\mathbf{z}_{ij})) \psi(\mathbf{v}_{ij}, \rho_{ij,b}(\mathbf{z}_{ij})) d\mathbf{v}_{ij} \right] dF(\mathbf{z}). \end{aligned} \quad (\text{F.4})$$

where the sum over (i, j) extends over all possible $\boldsymbol{\rho}_{ij}$ contained in the correlation vector $\boldsymbol{\rho}$. This should be put together with the first term of (F.2);

$$\frac{1}{\sqrt{n}} \int_S h(\boldsymbol{\rho}_b(\mathbf{z})) dG_n(\mathbf{z}) = \frac{1}{n} \sum_{t=1}^n h(\boldsymbol{\rho}_b(\mathbf{Z}_t)) - \int h(\boldsymbol{\rho}_b(\mathbf{z})) dF(\mathbf{z}). \quad (\text{F.5})$$

We have now reached an expression where one can easily apply the mixing central limit theorem, and an application of that theorem to the sum in (F.4) and (F.5) now yields asymptotic normality of $\sqrt{n}(T_{n,b} - T_b)$.

Using the same argument as in Lacal and Tjøstheim (2017) separately on the two main terms yields zero means. Moreover, using the definitions of G_n and $A_b(z)$ it can be seen from (F.4) and (F.5) that asymptotically

$$\sqrt{n}(T_{n,b} - T_b) \sim \int_S A_b(\mathbf{z}) dG_n(\mathbf{z}).$$

By analogy with the proof of Theorem 3.1 in Lacal and Tjøstheim (2017) it is seen that

$$\begin{aligned} \text{Var} \left(\int A_b(\mathbf{z}) dG_n(\mathbf{z}) \right) &\sim \int A_b^2(\mathbf{z}) dF(\mathbf{z}) + \frac{1}{n} \sum_{s \neq t} \int A_b(\mathbf{z}) A_b(\mathbf{u}) dF^{s,t}(\mathbf{z}, \mathbf{u}) \\ &\quad - n \int A_b(\mathbf{z}) A_b(\mathbf{u}) dF(\mathbf{z}) dF(\mathbf{u}) \\ &= C_n(A_b), \end{aligned}$$

which is the expression (F.1) stated in the theorem. The sum in the second term taken together with the last term above converges due to the mixing condition. Details of this argument are identical to those of the last part of the proof of Theorem 3.1 in Lacal and Tjøstheim (2019), as is the case of $\mathbf{b}_n \rightarrow \mathbf{0}$, completing the proof by referring to Proposition 6.3.9 of Brockwell and Davis (2006). ■

It should be carefully noted that it is not entirely obvious what the order of the term $C_n(A_b)$ is as $\mathbf{b} = (b_i, b_j) = (b, b) \rightarrow \mathbf{0}$ and $n \rightarrow \infty$. Looking at the expressions (F.4) and (F.5), it is the term (F.4) that causes potential trouble due to the presence of b_i^{-1} and b_j^{-1} in the two kernel function terms in the product kernel. These are cancelled out by the integration over (z_i, z_j) , however. This means that the true convergence rate for $T_{n,b} - T$ is $n^{-1/2}$. This may be somewhat surprising since the convergence rate of $\boldsymbol{\rho}_{n,b}(\mathbf{z}_{ij}) - \boldsymbol{\rho}(\mathbf{z}_{ij})$ is $(nb_1b_2)^{-1/2}$, but this is a well-known phenomenon for functionals of type $T_{n,b}$. In Lacal and Tjøstheim (2019, Remark 3.1) the convergence rate is changed from $(nb_1^3b_2^3)^{-1/2}$ to $(nb_1^2b_2^2)^{-1/2}$. These results are consistent with similar results for test functionals for independence in Joe (1989).

The variance in (F.5) can be evaluated in more detail, as has been done by Lacal and Tjøstheim (2019), in particular using the mixing inequality and Hall and Heyde (1980, Corollary A2). But as shown by Teräsvirta, Tjøstheim, and Granger (2010) and in numerous other papers, the asymptotic variance in the distribution of such functionals as $T_{n,b}$ cannot be expected to be accurate unless one has extremely large sample sizes.

The second alternative mentioned prior to stating Theorem F.2 is the alternative with $d_1 = d_2 = d_3 = 1$, but with correlations depending on three coordinates so that $\mathbf{z} = (z_1, z_2, z_3)$. This is easier in the sense that there are only three pairs of correlations, $\rho_{12}, \rho_{13}, \rho_{23}$ and in the above notation $\boldsymbol{\rho}(\mathbf{z}) = (\rho_{12}(\mathbf{z}), \rho_{13}(\mathbf{z}), \rho_{23}(\mathbf{z}))$. The partial correlation is

$$\alpha_{12|3}(\mathbf{z}) = \frac{\rho_{12}(\mathbf{z}) - \rho_{13}(\mathbf{z})\rho_{23}(\mathbf{z})}{\sqrt{1 - \rho_{13}^2(\mathbf{z})}\sqrt{1 - \rho_{23}^2(\mathbf{z})}},$$

which means that a natural choice of the function $h(\boldsymbol{\rho}(\mathbf{z}))$ might be $h(\boldsymbol{\rho}(\mathbf{z})) = \alpha_{12|3}^2(\mathbf{z})$. This is a three-parameter function of the parameters $\rho_{12}, \rho_{13}, \rho_{23}$, and an explicit expression for the gradient $\nabla_{\boldsymbol{\rho}} h(\boldsymbol{\rho})$ can easily be found. More generally again we define

$$T_{n,b} = \int_S h(\boldsymbol{\rho}_{b,n}(\mathbf{z})) dF_n(\mathbf{z}), \quad T_b = \int_S h(\boldsymbol{\rho}_b(\mathbf{z})) dF(\mathbf{z}) \quad \text{and} \quad T = \int_S h(\boldsymbol{\rho}(\mathbf{z})) dF(\mathbf{z}).$$

Theorem F.3. *Let the assumptions of Theorem F.2 be fulfilled and with a 3-coordinate $\boldsymbol{\rho}$ as defined above. Then, as $n \rightarrow \infty$,*

$$\sqrt{n} (C_n(A_b))^{-\frac{1}{2}} (T_{n,b} - T_b) \xrightarrow{d} \mathcal{N}(0, 1).$$

Here

$$A_b(\mathbf{z}) = \int_S \nabla_{\boldsymbol{\rho}} h(\boldsymbol{\rho}_b(\mathbf{v}))^T \mathbf{J}^{-1}(\boldsymbol{\rho}_b(\mathbf{v})) K_b(\mathbf{z} - \mathbf{v}) \mathbf{u}(\mathbf{z}, \boldsymbol{\rho}_b(\mathbf{v})) dF(\mathbf{v}) + h(\boldsymbol{\rho}_b(\mathbf{z})) 1_S(\mathbf{z})$$

and

$$C_n(A_b) = \int A_b^2(\mathbf{z}) dF(\mathbf{z}) + \frac{1}{n} \sum_{s \neq t} \int A_b(\mathbf{z}) A_b(\mathbf{u}) dF^{s,t}(\mathbf{z}, \mathbf{u}) - n \int A_b(\mathbf{z}) A_b(\mathbf{u}) dF(\mathbf{z}) dF(\mathbf{u})$$

with $F^{s,t}(\mathbf{z}, \mathbf{u})$ being the joint cumulative distribution function of $(\mathbf{Z}^s, \mathbf{Z}^t)$. Further, if $n \rightarrow \infty$ and $\mathbf{b}_n \rightarrow \mathbf{0}$, the relationship between them being as in Assumption 3 of Theorem 3.1 in the main article, then T_b can be replaced by T such that

$$\sqrt{n} (C_n(A_b))^{-\frac{1}{2}} (T_{n,b} - T) \xrightarrow{d} \mathcal{N}(0, 1).$$

Proof. The proof follows along the lines of the proof of Theorem F.2. Corresponding to equation (F.2),

$$T(\boldsymbol{\rho}_{b,n}) - T(\boldsymbol{\rho}_b) \sim \frac{1}{\sqrt{n}} \left\{ \int_S h(\boldsymbol{\rho}_b(\mathbf{z})) dG_n(\mathbf{z}) + \int_S \nabla_{\boldsymbol{\rho}} h(\boldsymbol{\rho}_b(\mathbf{z}))^T \mathbf{V}_n(\mathbf{z}) \right\} dF(\mathbf{z})$$

where $G_n(\mathbf{z}) = \sqrt{n}(F_n(\mathbf{z}) - F(\mathbf{z}))$ and

$$\begin{aligned} \mathbf{V}_n(\mathbf{z}) &= \sqrt{n}(\boldsymbol{\rho}_{n,b}(\mathbf{z}) - \boldsymbol{\rho}_b(\mathbf{z})) \\ &\sim \sqrt{n} \mathbf{J}_b^{-1}(\mathbf{z}) \left\{ \frac{1}{n} \sum_{t=1}^n K_b(\mathbf{Z}_t - \mathbf{z}) \mathbf{u}(\mathbf{Z}_t, \boldsymbol{\rho}_b(\mathbf{z})) - \int K_b(\mathbf{v} - \mathbf{z}) \mathbf{u}(\mathbf{v}, \boldsymbol{\rho}_b(\mathbf{z})) \psi(\mathbf{v}, \boldsymbol{\rho}_b(\mathbf{z})) d\mathbf{v} \right\}. \end{aligned}$$

Here K_b consists of a triple product of kernels $\prod_{i=1}^3 \frac{1}{b_i} K(\cdot/b_i)$, $\mathbf{u}(\mathbf{z}, \boldsymbol{\rho}_b(\mathbf{z})) = \nabla_{\boldsymbol{\rho}} \log \psi(\mathbf{z}, \boldsymbol{\rho}_b(\mathbf{z}))$ where ψ is the density function of a zero mean trivariate normal distribution having a covariance matrix equal to the correlation matrix and with off diagonal elements equal to the elements in $\boldsymbol{\rho}_b(\mathbf{z})$. This is further developed as in the proof of Theorem F.2, and the asymptotic normality of $T_{n,b} - T_b$ follows as before from the mixing assumptions and the proof is completed by applying Proposition 6.3.9 of Brockwell and Davis (2006). ■

The order of $C_n(A_b)$ is of interest as $\mathbf{b} \rightarrow \mathbf{0}$ and $n \rightarrow \infty$ since it determines the convergence rate of $T_{n,b} - T$. From Theorem 3.1 in the main article it follows that $\boldsymbol{\rho}_{n,b}(\mathbf{z})$ converges at a rate of $(nb^5)^{-1/2}$ to $\boldsymbol{\rho}(\mathbf{z})$ as $n \rightarrow \infty$ and $b_1 = b_2 = b_3 = b \rightarrow 0$. In the proof of Theorem F.3 one integrates over a triple kernel function in $C_n(A_b)$. This cancels out three b -s. so that one is left with a convergence rate of $(nb^2)^{-1/2}$ for $T_{n,b} - T$. This is slower than the convergence rate in Theorem F.3, but this is not unexpected since the faster rate in Theorem F.3 comes at the price of a 1-parameter pairwise simplification in the model structure.

G LOCAL ASYMPTOTIC POWER AND THE PITMAN CRITERION

Since the rates of convergence in Theorems F.2 and F.3 are $n^{-1/2}$ and $(nb^2)^{-1/2}$ respectively, the Pitman efficiency will be improved in the case of Theorem F.2 as compared to that of Theorem F.3.

We will follow the approach used by Wang and Hong (2018), in the proof of their Theorem 2 to derive the results on local asymptotic power. We will explicitly use the results in Theorems F.2 and F.3 with $h(\boldsymbol{\alpha}_{n,b}(\mathbf{z})) = \|\boldsymbol{\alpha}_{n,b}(\mathbf{z})\|^2$, so that $T_{n,b} = \int_S \|\boldsymbol{\alpha}_{n,b}(\mathbf{z})\|^2 dF_n(\mathbf{z})$. The null hypothesis of conditional independence can then be phrased as

$$H_0 : \|\boldsymbol{\alpha}(\mathbf{z})\|^2 = 0 \text{ for } \mathbf{z} \text{ almost everywhere in } S.$$

against the Pitman alternative

$$H[a_n] : \|\boldsymbol{\alpha}(\mathbf{z})\|^2 = a_n \beta(\mathbf{z})$$

where a_n is a sequence that tends to zero at a certain rate, determining the asymptotic rate at which there is still power in testing H_0 , and $\beta(\mathbf{z}) = \beta(\rho(\mathbf{z}))$ is a positive function of \mathbf{z} such that $\int_S \beta(\mathbf{z}) dF(\mathbf{z}) \stackrel{\text{def}}{=} \gamma < \infty$. Here $\beta(\mathbf{z})$ is the result of taking an arbitrary correlation vector $\rho(\mathbf{z}) = \rho_H(\mathbf{z})$ under the alternative hypothesis such that $\boldsymbol{\alpha}(\rho_H(\mathbf{z})) \neq \mathbf{0}$ and $\beta(\mathbf{z})$ is the Froebenius norm squared of this $\boldsymbol{\alpha}$. Further, let $C_{H,n,b}(A_b)$ be the $C_{n,b}$ -function formed as in Theorem F.2 from $\rho_H(\mathbf{z})$. It converges towards $C_H(\mathbf{A}_b)$ as depicted in the mixing expression of the proof of Theorem F.2. See also Theorem 3.2 of Lacal and Tjøstheim (2019) which is formulated without using independence.

Theorem G.1. *Let the conditions be as in Theorem F.2 and let $H[a_n]$ be as above with $a_n = n^{-1/2}$. Let $C_H(A)$ be the limit of $C_n(A_b)$ as $n \rightarrow \infty$ and $\mathbf{b} \rightarrow \mathbf{0}$ at the rates specified in the assumptions of Theorem F.2. (This limit exists using the mixing expressions of the proof of Theorem F.2). Then the power of $n^{1/2}(C_{H,n,b}(A_b))^{-1/2}(T_{n,b} - T)$ ($T = 0$ under our H_0 , but the procedure below can also be carried through with a general $\boldsymbol{\alpha}$ under H_0) is given by $\mathbb{P}[T_{n,b} \geq z_\varepsilon | H[a_n]] \rightarrow 1 - \Phi(z_\varepsilon - \gamma/[C_H(A)]^{1/2})$ as $n \rightarrow \infty, \mathbf{b} \rightarrow \mathbf{0}$, where $\Phi(\cdot)$ is the cumulative standard normal distribution function and z_ε is the one-sided critical value of $\mathcal{N}(0, 1)$ at significance level ε .*

Proof. The test statistic can be written as

$$\begin{aligned} T_{n,b} &= \int_S \|\boldsymbol{\alpha}(\mathbf{z})_{n,b}\|^2 dF_n(\mathbf{z}) \\ &= \int_S \|\boldsymbol{\alpha}(\mathbf{z})_{n,b}\|^2 dF_n(\mathbf{z}) - a_n \int_S \beta(\mathbf{z}) dF(\mathbf{z}) + a_n \int_S \beta(\mathbf{z}) dF(\mathbf{z}). \end{aligned}$$

Analogously to Wang and Hong (2018), arguing as in the proof of Theorem F.2 and using Proposition 6.3.9 of Brockwell and Davis (2006), as $b_n, a_n \rightarrow 0$,

$$\sqrt{n}[C_{H,n,b}(A_b)]^{-1/2} \left\{ \int_S \|\boldsymbol{\alpha}(\mathbf{z})_{n,b}\|^2 dF_n(\mathbf{z}) - a_n \int_S \beta(\mathbf{z}) dF(\mathbf{z}) \right\} \xrightarrow{d} \mathcal{N}(0, 1).$$

and

$$\sqrt{n}[C_{H,n,b}(A_b)]^{-1/2}a_n \int_S \beta(\mathbf{z}) dF(\mathbf{z}) \rightarrow \gamma/[C_H(A)]^{1/2}$$

and this completes the proof. ■

The alternative situation that is covered in Theorem F.3 can now be stated and proved analogously. In this case $b^2 C_{H,n,b}$ converges towards $C'(\mathbf{A})$, say, where b is the common value of the bandwidths (b_1, b_2, b_3) .

Theorem G.2. *Let the conditions be as in Theorem G.1 and let $H[a_n]$ be as above with $a_n = (nb^2)^{-1/2}$, with b the common value of the three bandwidths (b_1, b_2, b_3) . Then, letting $C'_H(A)$ be the limit of $b^2 C_{H,n,b}(A_b)$ as $n \rightarrow \infty$ and $b \rightarrow 0$ at the rates specified in the assumptions of Theorems F.2 and F.3. (This limit exists using the mixing expressions of the proof of Theorem F.2). Then the power of $(nb^2)^{1/2}(C_{H,n,b}(A_b))^{-1/2}(T_{n,b} - T)$ ($T = 0$ under our H_0 , but the procedure can also be carried through with a general α under H_0 .) is given by $\mathbb{P}[T_{n,b} \geq z_\varepsilon | H[a_n]] \rightarrow 1 - \Phi(z_\varepsilon - \gamma/[C_H(A)']^{1/2})$ as $n \rightarrow \infty, b \rightarrow 0$, where $\Phi(\cdot)$ is the cumulative standard normal distribution function and z_ε is the one-sided critical value of $\mathcal{N}(0, 1)$ at significance level ε .*

We conclude that the testing procedure can pick up alternatives that converges at the rate $(nb^2)^{-1/2}$ towards the hypothesis of conditional independence as formulated under the alternative (i) in Section A.3.

For a further brief discussion, see the paragraph following eqs. (27) and (28) in Section 5.2 in the main article.

H BOOTSTRAPPING THE TEST STATISTIC

The main principle of this section is that we replace F_n -quantities with F_n^* -quantities and F -quantities with F_n -quantities, the latter being known under the bootstrap operation. Let

$$T_{n,b}^* = \int_S h(\rho_{b,n}^*(\mathbf{z})) dF_n^*(\mathbf{z}).$$

We treat the pairwise model case corresponding to Theorem F.2 first.

Theorem H.1. *Let the assumptions of Theorem E.2 be fulfilled. Then as $n \rightarrow \infty$*

$$\sqrt{n}[C_n(A_b)]^{-1/2}(T_{n,b}^* - T_{n,b}) \xrightarrow{d_{\mathbb{P}^*}} \mathcal{N}(0, 1) \text{prob} - \mathbb{P}.$$

Here

$$A_b(\mathbf{z}) = \int_S \left[\sum_{i,j} \frac{\partial}{\partial \rho_{ij,n,b}} h(\rho_{n,b}(\mathbf{v})) J_{ij}^{-1}(\rho_{ij,n,b}(\mathbf{v}_{ij})) K_b(\mathbf{z}_{ij} - \mathbf{v}_{ij}) u_{ij}(\mathbf{z}_{ij}, \rho_{ij,n,b}(\mathbf{v}_{ij})) \right] dF(\mathbf{v}) \\ + h(\rho_{ij,n,b}(\mathbf{z})) 1_S(\mathbf{z}),$$

and

$$C_n(A_b) = \int A_b^2(\mathbf{z}) dF_n(\mathbf{z}) + \frac{1}{n} \sum_{s \neq t} \int A_b(\mathbf{z}) A_b(\mathbf{u}) dF_n^{s,t}(\mathbf{z}, \mathbf{u}) \\ - n \int A_b(\mathbf{z}) A_b(\mathbf{u}) dF_n(\mathbf{z}) dF_n(\mathbf{u}), \quad (\text{H.1})$$

with $F_n^{s,t}(\mathbf{z}, \mathbf{u})$ being the empirical joint cumulative distribution function of $(\mathbf{Z}_s, \mathbf{Z}_t)$. Note that the sum in the second term in the definition of $C_n(A_b)$ taken together with the third term converges due to the mixing assumption.

Proof. As in the beginning of the proof of Theorem 4.4, part B, of the supplement of Lacal and Tjøstheim (2017), one has that asymptotically,

$$T_{n,b}^* - T_{n,b} \sim \frac{1}{\sqrt{n}} \left\{ \int_S h(\boldsymbol{\rho}_b(\mathbf{z})) dG_n^*(\mathbf{z}) + \int_S \nabla_{\boldsymbol{\rho}} h(\boldsymbol{\rho}_b(\mathbf{z}))^T \tilde{V}_n(\mathbf{z}) dF_n(\mathbf{z}) \right\}, \quad (\text{H.2})$$

where

$$G_n(\mathbf{z}) = \sqrt{n}(F_n^*(\mathbf{z}) - F_n(\mathbf{z})) \quad \text{and} \quad \tilde{V}_n(\mathbf{z}) = \sqrt{n}((\boldsymbol{\rho}_{n,b}^*(\mathbf{z}) - \boldsymbol{\rho}_{n,b}(\mathbf{z})))$$

with F_n^* being the empirical cumulative function for the bootstrapped time series, and where again $\boldsymbol{\rho}$ is the vector consisting of all possible pairs ρ_{ij} . This rewriting of $T_{n,b}^* - T_{n,b}$ is the first step in obtaining an expression in terms over \mathbf{Z}_t^* , $t = 1, \dots, n$ so that a mixing type CLT can be used as done in the paper by Gonçalves and White (2002). We proceed to the pairwise analysis.

Let (i, j) be a fixed pair and $\rho_{ij,n,b}^*(z_i, z_j)$ be the corresponding estimated bootstrap local correlation for the pair (i, j) chosen from the vector $\boldsymbol{\rho}_{n,b}^*$. Here the pair of coordinates (z_i, z_j) corresponds to the bivariate \mathbf{x} in Lacal and Tjøstheim (2017, proof of Theorem 3.1). To ease notation and the comparison with Lacal and Tjøstheim (2017, 2019), we again write $\mathbf{z}_{ij} = (z_i, z_j)$. We will also omit some indices or parts of indices when it is clear what is meant. Finally, we omit the area of integration S . We are using the pairwise log likelihood given by

$$L_{ij,n}^* = \frac{1}{n} \sum_{t=1}^n K_b(\mathbf{Z}_{ij,t}^* - \mathbf{z}_{ij}) \log \psi(\mathbf{Z}_{ij,t}^*, \rho_{ij,n,b}(\mathbf{z}_{ij})) - \int K_b(\mathbf{v}_{ij} - \mathbf{z}_{ij}) \psi_{ij}(\mathbf{v}_{ij}, \rho_{ij,n,b}(\mathbf{z}_{ij})) d\mathbf{v}_{ij},$$

where $K_b = \frac{1}{b_1} K_1(\cdot/b_1) \frac{1}{b_2} K_2(\cdot/b_2)$ is a product kernel function and $\mathbf{Z}_{ij,t}^* = (Z_{i,t}^*, Z_{j,t}^*)$ is the t -th pair of bootstrap observations. Estimates $\rho_{ij,n,b}^*(\mathbf{z}_{ij})$ are obtained by differentiation of the log likelihood and by introducing the score function U_{ij}^* and information function I_{ij}^* as follows

$$\begin{aligned} U_{ij}^*(\mathbf{z}_{ij}, \rho_{ij,n}) &= \frac{\partial}{\partial \rho_{ij,n}} L_{ij,n}^*(\mathbf{z}_{ij}, \rho_{ij,n}) \\ &= \frac{1}{n} \sum_{t=1}^n K_b(\mathbf{Z}_{ij,t}^* - \mathbf{z}_{ij}) u_{ij}(\mathbf{Z}_{ij,t}^*, \rho_{ij,n}) - \int K_b(\mathbf{v}_{ij} - \mathbf{z}_{ij}) u_{ij}(\mathbf{v}_{ij}, \rho_{ij,n}) \psi(\mathbf{v}_{ij}, \rho_{ij,n}) d\mathbf{v}_{ij} \end{aligned}$$

where $u_{ij}(\mathbf{z}_{ij}, \rho_{ij,n}) = \frac{\partial}{\partial \rho_{ij,n}} \log \psi(\mathbf{z}_{ij}, \rho_{ij,n})$. Similarly,

$$I_{ij}^*(\mathbf{z}_{ij}, \rho_{ij,n}) = \frac{\partial^2}{\partial \rho_{ij,n}^2} L_{ij,n}^*(\mathbf{z}_{ij}, \rho_{ij,n}) \quad \text{and} \quad J_{ij,n,b}(\mathbf{z}_{ij}) = \mathbb{E}^* \left(-I_{ij}^*(\mathbf{z}_{ij}, \rho_{ij,n}) \right),$$

where $\mathbb{E}^*(\cdot)$ is the mean value under the measure \mathbb{P}_{ω}^* , and where Lacal and Tjøstheim (2017, proof of Theorem 4.4) may be consulted for more details in the derivation and a more explicit analog expression for $I_{ij}^*(\mathbf{z}_{ij}, \rho_{ij,n})$ for the case considered there. In contradistinction to Lacal and Tjøstheim (2017), $J_{ij,n,b}$ is not singular as $n \rightarrow \infty, \mathbf{b} \rightarrow 0$. In fact, it can

be shown that, cf. Otneim and Tjøstheim (2017, Theorem 2), as $n \rightarrow \infty$ and $\mathbf{b} \rightarrow \mathbf{0}$, $J_{ij,n,b}(\mathbf{z}_{ij}) \xrightarrow{\mathbb{P}} u(\mathbf{z}_{ij}, \rho_{ij})\psi(\mathbf{z}_{ij}, \rho_{ij}(\mathbf{z}_{ij}))$.

The vector $\tilde{V}_n(\mathbf{z}) = \sqrt{n}(\boldsymbol{\rho}_{n,b}^*(\mathbf{z}) - \boldsymbol{\rho}_{n,b}(\mathbf{z}))$ has the dimension of $\boldsymbol{\rho}(\mathbf{z})$ and consists of components $\{\rho_{ij,n,b}^*(\mathbf{z}_{ij} - \rho_{ij,n,b}(\mathbf{z}_{ij}))\}$ where i and j in the most general case run over all possible pairs formed by the components of $\boldsymbol{\rho}(\mathbf{z})$. To be able to use the mixing CLT analogous to Gonçalves and White (2002) we follow Lacal and Tjøstheim (2017) and express this more explicitly in terms of the variables \mathbf{Z}_t^* , $t = 1, \dots, n$. By Taylor expansion as in the proof of Theorem F.2, and omitting the index ij on ρ ,

$$\begin{aligned} 0 &= U_{ij}^*(\mathbf{z}_{ij}, \rho_{n,b}^*(\mathbf{z}_{ij})) \\ &\sim U_{ij}^*(\mathbf{z}_{ij}, \rho_{n,b}(\mathbf{z}_{ij})) + I_{ij}^*(\mathbf{z}_{ij}, \rho_{n,b}(\mathbf{z}_{ij}))(\rho_{n,b}^*(\mathbf{z}_{ij}) - \rho_{n,b}(\mathbf{z}_{ij})) \\ &\sim U_{ij}^*(\mathbf{z}_{ij}, \rho_{n,b}(\mathbf{z}_{ij})) - J_{ij,n,b}(\mathbf{z}_{ij})(\rho_{n,b}^*(\mathbf{z}_{ij}) - \rho_{n,b}(\mathbf{z}_{ij})) \end{aligned}$$

where we have used that $E^*(\cdot)$ is the mean value under the measure \mathbb{P}_ω^* and again following the reasoning of the proof of Theorem F.2, this implies that

$$\begin{aligned} \rho_{n,b}^*(\mathbf{z}_{ij}) - \rho_{n,b}(\mathbf{z}_{ij}) &\sim J_{ij,n,b}^{-1}(\mathbf{z}_{ij})U_{ij}^*(\mathbf{z}_{ij}, \rho_{n,b}(\mathbf{z}_{ij})) \\ &= J_{ij,n,b}^{-1}(\mathbf{z}_{ij}) \left[\frac{1}{n} \sum_{t=1}^n K_b(\mathbf{Z}_{ij,t}^* - \mathbf{z}_{ij})u_{ij}(\mathbf{Z}_{ij,t}^*, \rho_{n,b}(\mathbf{z}_{ij})) \right. \\ &\quad \left. - \int K_b(\mathbf{v}_{ij} - \mathbf{z}_{ij})u_{ij}(\mathbf{v}_{ij}, \rho_{n,b}(\mathbf{z}_{ij}))\psi(\mathbf{v}_{ij}, \rho_{n,b}(\mathbf{z}_{ij}))d\mathbf{v}_{ij} \right]. \quad (\text{H.3}) \end{aligned}$$

The second term in (H.2) can then be expressed as

$$\begin{aligned} \int_S \sum_{i,j} \frac{\partial}{\partial \rho_{ij,n,b}} h(\rho_{n,b}(\mathbf{z})) \times J_{ij,n,b}^{-1}(\mathbf{z}_{ij}) \left[\frac{1}{n} \sum_{t=1}^n K_b(\mathbf{Z}_{ij,t}^* - \mathbf{z}_{ij})u_{ij}(\mathbf{Z}_{ij,t}^*, \rho_{n,b}(\mathbf{z}_{ij})) \right. \\ \left. - \int K_b(\mathbf{v}_{ij} - \mathbf{z}_{ij})u_{ij}(\mathbf{v}_{ij}, \rho_{n,b}(\mathbf{z}_{ij}))\psi(\mathbf{v}_{ij}, \rho_{n,b}(\mathbf{z}_{ij}))d\mathbf{v}_{ij} \right] dF_n(\mathbf{z}), \quad (\text{H.4}) \end{aligned}$$

where the sum over (i, j) extends over all ρ_{ij} contained in the ρ -vector. This should be put together with the first term of (H.2),

$$\frac{1}{\sqrt{n}} \int_S h(\rho_{n,b}(\mathbf{z})) dG_n^*(\mathbf{z}) = \frac{1}{n} \sum_{t=1}^n h(\rho_{n,b}(\mathbf{Z}_t^*)) - \int h(\rho_{n,b}(\mathbf{z})) dF_n(\mathbf{z}) \quad (\text{H.5})$$

and we have thus reached an expression where we can apply the mixing CLT as in Gonçalves and White (2002) and an application of their result, quoted as Theorems C.1 and E.1 in the supplementary materials of Lacal and Tjøstheim (2017) and Lacal and Tjøstheim (2019) respectively, to the expressions in (H.4) and (H.5). This yields asymptotic normality of $\sqrt{n}(T_{n,b}^* - T_{b,n})$.

The rest of the proof now follows along the lines of the end of the proof of Theorem F.2, including the use of Proposition 6.3.9 of Brockwell and Davis (2006). ■

An analogue of Theorem F.3 can now be phrased and proved similarly.

Theorem H.2. *Let the assumptions of Theorem E.2 be fulfilled. Then, as $n \rightarrow \infty$ and $\mathbf{b} \rightarrow \mathbf{0}$*

$$\sqrt{n}(C_n(A_b))^{-\frac{1}{2}} \left((T_{n,b}^* - T_{n,b}) \xrightarrow{d_{\mathbb{P}_\omega^*}} \mathcal{N}(0, 1) \text{prob} - \mathbb{P} \right).$$

Here

$$A_b(\mathbf{z}) = \int_S \nabla_{\rho} h(\rho_{n,b}(\mathbf{v}))^T \mathbf{J}^{-1}(\rho_{n,b}, \mathbf{v}) K_b(\mathbf{z} - \mathbf{v}) \mathbf{u}(\mathbf{z}, \rho_{n,b}(\mathbf{v})) dF(\mathbf{v}) + h(\rho_{n,b}(\mathbf{z})) 1_S(\mathbf{z})$$

and

$$\begin{aligned} C_n(A_b) = & \int A_b^2(\mathbf{z}) dF_n(\mathbf{z}) + \frac{1}{n} \sum_{s \neq t} \int A_b(\mathbf{z}) A_b(\mathbf{u}) dF_n^{s,t}(\mathbf{z}, \mathbf{u}) \\ & - n \int A_b(\mathbf{z}) A_b(\mathbf{u}) dF_n(\mathbf{z}) dF_n(\mathbf{u}), \end{aligned}$$

with $F_n^{s,t}(\mathbf{z}, \mathbf{u})$ being the empirical joint cumulative distribution function of $(\mathbf{Z}_s, \mathbf{Z}_t)$.

I THE LOCAL BOOTSTRAP

The limiting results for $\alpha_n(\mathbf{z})$ (or $\boldsymbol{\alpha}_n(\mathbf{z})$ in the matrix case) and $T_{n,b}$, and the corresponding bootstrap results can be used to construct asymptotic and bootstrap confidence intervals of T_b and T . In that sense they give an indication of whether $\alpha(\mathbf{z}) = 0$ locally or $\alpha(\mathbf{z}) \equiv 0$ globally respectively. Such confidence intervals, however, cannot in general be used to test for $\alpha(\mathbf{z}) \equiv 0$, i.e. conditional independence, with a given level of the test. To be able to do that one must devise a test statistic that is constructed with a probability structure as under the null hypothesis of conditional independence. It is difficult to obtain such a test based on asymptotic arguments: As mentioned before the asymptotic distribution of test functionals such as $T_{n,b}$ is known to be inaccurate except for extremely large sample sizes. It is also difficult to build the probability null structure into the scaling factor of $T_{n,b}$ since this would amount to building into that structure certain relationships between pairwise local correlations.

An alternative that has been used in the testing theory of conditional independence is the so-called local bootstrap. This is a device that uses testing based on bootstrapping from kernel density estimates. It was used for this purpose by Paparoditis and Politis (2000) and much more extensively by Su and White (2008). It has also been used (without proof) by Wang and Hong (2018) and by Huang (2010). We have used a variation of the local bootstrap in our test of conditional independence with good results in accuracy of level and in power. The primary purpose of the present paper is not to prove in detail properties of the local bootstrap in local Gaussian testing, but we will nevertheless indicate theoretically why a local type bootstrap argument makes sense, and why it can be expected to, and does, work also in our case. To put the local type bootstrap into our test functional context we need first to state some facts and properties of the local bootstrap as used in the literature. Note that our functional is in a sense analogous to that used by Wang and Hong (2018).

In our notation $\mathbf{Z}_{3,t}$ is the basic d_3 -dimensional process corresponding to X_t in Wang and Hong (2018). Under H_0 we assume that $\mathbf{Z}_{1,t}$ and $\mathbf{Z}_{2,t}$ of dimensions d_1 and d_2 are conditionally independent given $\mathbf{Z}_{3,t}$, where $\mathbf{Z}_{1,t}$ and $\mathbf{Z}_{2,t}$ correspond to Y_t and Z_t in Wang and Hong (2018).

The first step consists in constructing a bootstrap realization $\{\mathbf{Z}_{3,t}^*\}$ of $\{\mathbf{Z}_{3,t}\}$ by sampling from the kernel density estimate $\tilde{f}(\mathbf{z}_3) = \frac{1}{n} \sum_{t=1}^n L_b(\mathbf{z}_3 - \mathbf{Z}_{3,t})$, where L_b is a d_3 dimensional

product kernel function with factors $L_{b_i} = \frac{1}{b_i} L(\cdot/b_i)$. The bootstrap realizations of $\{\mathbf{Z}_{1,t}^*\}$ and $\{\mathbf{Z}_{2,t}^*\}$ are then created independently of each other by sampling from the conditional kernel estimators $\tilde{f}(z_1|\mathbf{Z}_{3,t}^*) = \sum_s L_b(\mathbf{Z}_{1,s} - z_1) L_b(\mathbf{Z}_{3,s} - \mathbf{Z}_{3,t}^*) / \sum_s L_b(\mathbf{Z}_{3,s} - \mathbf{Z}_{3,t}^*)$ and $\tilde{f}(z_2|\mathbf{Z}_{3,t}^*) = \sum_s L_b(\mathbf{Z}_{2,s} - z_2) L_b(\mathbf{Z}_{3,s} - \mathbf{Z}_{3,t}^*) / \sum_s L_b(\mathbf{Z}_{3,s} - \mathbf{Z}_{3,t}^*)$, respectively.

The bootstrap test statistic T_n^* (which is a Hellinger type statistic in the Su-White case. can then be constructed from these bootstrap observations $\{\mathbf{Z}_{1,t}^*, \mathbf{Z}_{2,t}^*, \mathbf{Z}_{3,t}^*\}$ using another kernel function K and a bandwidth h to construct density estimates $\hat{f}^*(z_3)$, $\hat{f}^*(z_2, z_3)$, $\hat{f}^*(z_1, z_3)$ and $\hat{f}^*(z_1, z_2, z_3)$ which is inserted in the Hellinger functional to obtain a bootstrap value T_n^* of this statistic. It can then be shown (Su and White 2008) that properly normalized $T_n^* \xrightarrow{d} \mathcal{N}(0, 1)$.

Further, if \mathbb{P}^* denote the probability conditional on the sample $\{\mathbf{Z}_t^*\}$, then the level- ε critical values \tilde{c}_ε are computed as approximate solutions of $\mathbb{P}^*(T_n^* > \tilde{c}_\varepsilon) = \varepsilon$ and the bootstrap p -value is given by $p^* = B^{-1} \sum_{j=1}^B 1(T_{n_j}^* > T_n)$, where B is the number of bootstrap realization. To integrate this procedure into our context is fairly straightforward. The first step is to generate local type bootstrap observations $\{\mathbf{Z}_{1,t}^*, \mathbf{Z}_{2,t}^*, \mathbf{Z}_{3,t}^*\}$. The bootstrap realization $\{\mathbf{Z}_{3,t}^*\}$ can be created using the ordinary bootstrap on $\{\mathbf{Z}_{3,t}\}$, cf. Paparoditis and Politis (2000), or the kernel smoothed bootstrap as in Su and White (2008). If the dimension of $\mathbf{Z}_{3,t}$ is large, a smoothed bootstrap realization can be created by sampling from the local Gaussian density estimate of Otneim and Tjøstheim (2017). In the next step in creating the conditional bootstrap realizations $\{\mathbf{Z}_{2,t}^*\}$ and $\{\mathbf{Z}_{1,t}^*\}$ one can use the smoothed conditional kernel estimate of Su and White (2008), or if the dimensions are large, to avoid the curse of dimensionality, one can sample from the conditional local Gaussian density estimates of Otneim and Tjøstheim (2018) using, say, acceptance reject sampling. Again it should be noted that the avoidance of the kernel estimate comes at the cost of a more restrictive model.

Creating bootstrap realizations by the above scheme is quite time consuming since for each $\mathbf{Z}_{3,t}^*$, drawings of $\mathbf{Z}_{1,t}^*$ and $\mathbf{Z}_{2,t}^*$ have to be done, and this has to be repeated over the number of bootstrap realizations. We have devised a simplified and more efficient procedure for this as described in the main article.

Once a bootstrap realization $\{\mathbf{Z}_{1,t}^*, \mathbf{Z}_{2,t}^*, \mathbf{Z}_{3,t}^*\}$ is created, following Su and White (2008), these realizations are now taken as the original observations in the local likelihood estimation procedure. For a fixed bandwidth b the population values, following the local Gaussian approach, can be taken as the solution of the integral equation (2) in Tjøstheim and Hufthammer (2013); see also corresponding equations in e.g. Otneim and Tjøstheim (2017, 2018) and eq. (19) in the main article. Note that in the present paper we have largely operated with two alternatives, the general vector case with inference via the pairwise local likelihood, or the the purely scalar case with three coordinates for the local Gaussian correlation. In both cases we have normalized the processes so that they have standard normal marginals.

In the $\{\mathbf{Z}_{1,t}^*, \mathbf{Z}_{2,t}^*, \mathbf{Z}_{3,t}^*\}$ -framework we obtain an estimate $\hat{\boldsymbol{\rho}}^*$ of the local Gaussian correlation vector conditional on the bootstrap realizations. An asymptotic theory of this estimate can now be established using the method of proof in Theorem 2 in Otneim and Tjøstheim (2017), and subsequently an asymptotic theory for the test statistic T_n^* using the method of proof in Theorem 4.4 in Lacal and Tjøstheim (2017) and Theorem D.4 in Lacal and Tjøstheim (2019). See also the formulation of such proofs in the present paper, theorems H.1 and H.2. The scaling factor $C_n(A_b)$ is of the same type as in the formulation of those results with

obvious notational changes. Moreover under the present null hypotheses $T = 0$, it may be conjectured that it leads to the asymptotic results that the scaled $T_n^* \xrightarrow{d_{P^*}} \mathcal{N}(0, 1) \text{prob} - \mathbb{P}$.

J TWO ADDITIONAL EXAMPLES OF ESTIMATION

We will in this section provide two more demonstrations on how the estimated LGPC can be used to reveal nonlinear departures from conditional independence and to measure that conditional dependence. The ordinary partial correlation coefficient fully characterizes conditional dependence between jointly Gaussian variables. It is not hard, though, to construct examples in which strong nonlinear relationships remain completely undetected by the partial correlation. Consider the following structural equation;

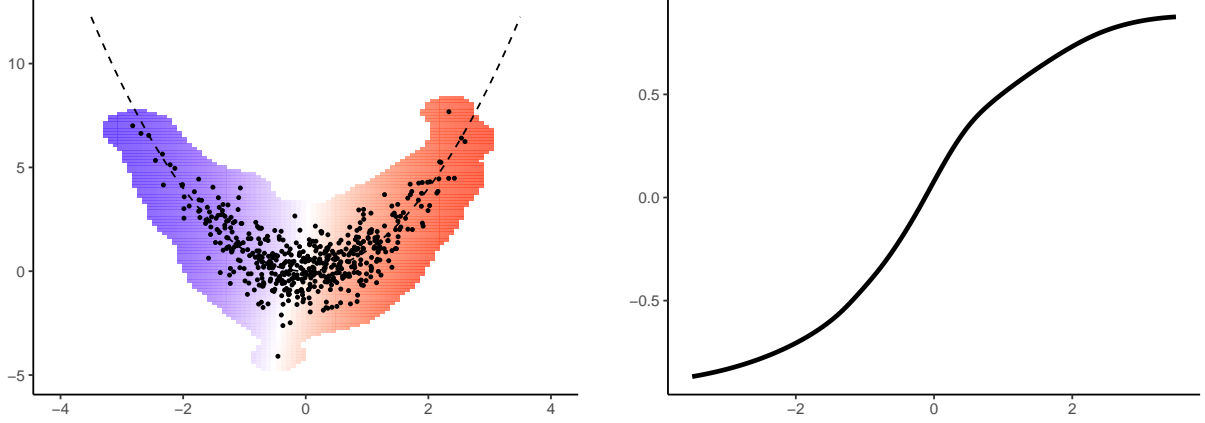
$$X_2 = X_1^2 + X_3, \tag{J.1}$$

where we observe all components in (X_1, X_2, X_3) . There is, obviously, a strong dependence between X_1 and X_2 , and furthermore, this dependence is deterministic when conditioning on $X_3 = x_3$. It is, however, well known that X_1 and $X_1^2 + x_3$ are uncorrelated if $E(X_1) = E(X_1^3) = 0$, but the LGPC easily reveals conditional dependence between X_1 and X_2 in this case. Let us generate $n = 500$ independent observations each from $X_1 \sim \mathcal{N}(0, 1)$ and $X_3 \sim \mathcal{N}(0, 1)$ and calculate X_2 by (J.1). The sample partial correlation between X_1 and X_2 is -0.037 in this case, but the LGPC, which in this case is calculated using the simplified pairwise structure defined in Section 3.1.2, displayed in Figure 1a, indicates strong departures from conditional independence on the plane defined by $X_3 = 0$. Indeed, we identify a region of *negative conditional relationship* on the left hand side of the plot, which makes good sense because small values of X_2 is typically observed together with large values of X_1 and vice versa. The opposite phenomenon is clearly visible in the right hand half of the plot. In Figure 1b we have plotted the estimated LGPC along the curve $X_2 = X_1^2$. The full trivariate fit presented in Section 3.1.1 gives a very similar picture.

This points to an important feature of the LGPC, as we also saw in one of the simulated examples in the main article: 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.

Let us briefly demonstrate a practical implementation of the LGPC in another situation involving real data. As in the empirical example in the main article, we look at Granger causality.

One fairly recent method for testing for Granger causality is based on the maximal conditional correlation and was introduced by Huang (2010) and later extended to the time series case by Cheng and Huang (2012). The latter authors use this test to examine whether trading volume Granger caused index value, or vice versa, on the S&P 500 stock index during the first decade of this century. Using the daily price series $\{P_t\}$ and volume $\{V_t\}$ they define the log-differenced series



(a) Estimated local Gaussian partial correlation between X_1 and X_2 given $X_3 = 0$ in the structural model $X_2 = X_1^2 + X_3$. The color scale goes from -1 (dark blue) to $+1$ (dark red).

(b) The estimated LGPC along the curve $X_2 = X_1^2$.

Figure 1: The structural model.

$$R_t = 100 \log \left(\frac{P_t}{P_{t-1}} \right) \quad \text{and} \quad V_t^* = \log \left(\frac{V_t}{V_{t-1}} \right).$$

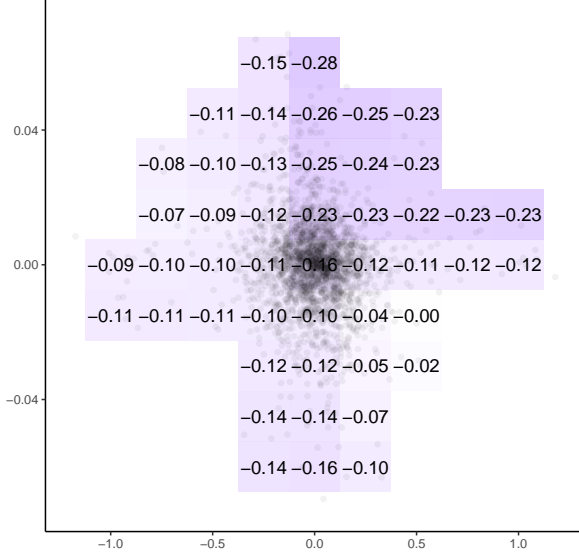
Further, they denote by $R_{t-1} \not\Rightarrow V_t^*$ the Granger non-causality from $\{R_t\}$ to $\{V_t^*\}$, which they analyze by putting the following null hypothesis to the test:

$$H_0 : V_t^* \perp R_{t-1} \mid V_{t-1}^*. \quad (\text{J.2})$$

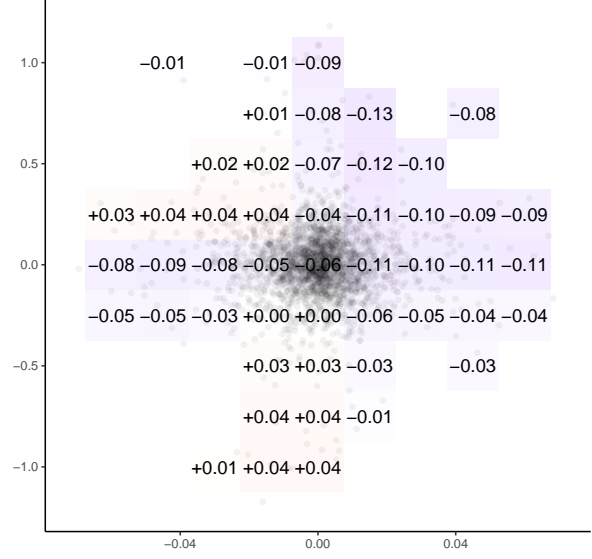
Granger non-causality in the opposite direction, $V_{t-1}^* \not\Rightarrow R_t$ is formulated and tested for in the obvious way.

A linear test rejects $R_{t-1} \not\Rightarrow V_t^*$, but not $V_{t-1}^* \not\Rightarrow R_t$. The nonparametric test by Cheng and Huang (2012) rejects both, leading to the natural conclusion that there are nonlinear relationships in the latter situation that remain unseen by linear models. What is it, though, in this particular data set, that leads to such results? Estimates of the local partial Gaussian correlation provide some clues towards answering that question.

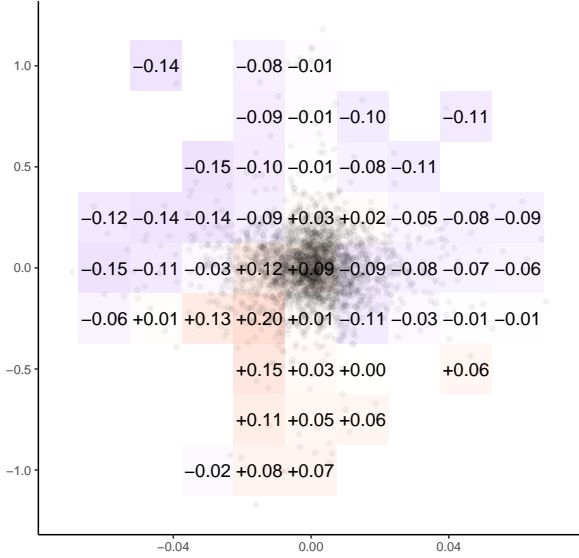
We obtain daily observations on price and trading volume on the S&P 500 index from January 1st 2000 until December 31st 2009 from Yahoo Finance (2018), and plot the estimated trivariate LGPC between V_t^* and R_{t-1} on the plane defined by $V_{t-1}^* = 0$ in Figure 2a. Departures from $\alpha(V_t^*, R_{t-1}) \equiv 0$ provide evidence against $R_{t-1} \not\Rightarrow V_t^*$. Similarly, we plot the estimated LGPC between R_t and V_{t-1}^* on the plane $R_{t-1} = 0$ in Figure 2b in which departures from $\alpha(R_t, V_{t-1}^*) \equiv 0$ provide evidence against $V_{t-1}^* \not\Rightarrow R_t$. We have used a proportionality constant of $c = 3.5$ to calculate the estimates (see discussion on bandwidth selection in Section 4 of the main article). The observations can be seen in the background of both plots.



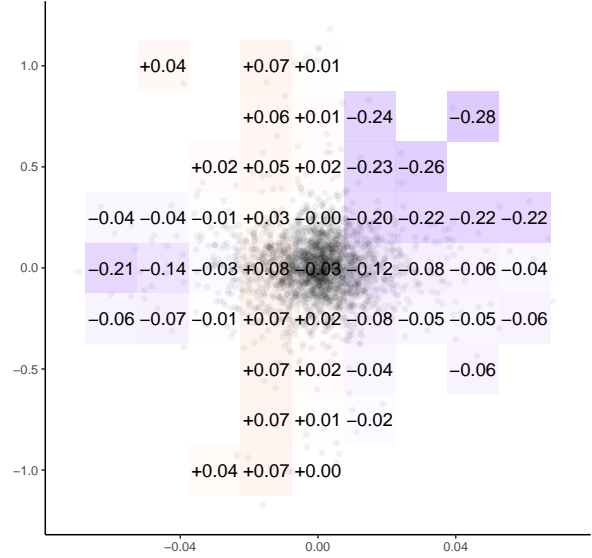
(a) Local partial dependence map of V_t^* vs. R_{t-1} on the plane defined by $V_{t-1}^* = 0$



(b) Local partial dependence map of R_t vs. V_{t-1}^* on the plane defined by $R_{t-1} = 0$



(c) Local partial dependence map of R_t vs. V_{t-1}^* on the plane defined by $R_{t-1} = -0.05$



(d) Local partial dependence map of R_t vs. V_{t-1}^* on the plane defined by $R_{t-1} = 0.05$

Figure 2: Local partial dependence maps for the S&P 500 data

The differences between Figures 2a and 2b are subtle, but important. In Figure 2a, $\hat{\alpha}(V_t^*, R_{t-1})$ is mostly negative, especially in the data rich portions of the sample space (other values of the conditioning variable V_{t-1}^* than zero give similar pictures). Indeed, the *global* partial correlation in this situation is $\hat{\alpha}_{\text{glob}} = -0.086$, which is significantly different from zero. The global partial correlation in the second situation is very small though: $\hat{\alpha}_{\text{glob}} = -0.0018$, but Figure 2b still reveals departures from conditional independence of similar magnitudes as in Figure 2a. The difference is that the estimated LGPC is positive (primarily in the second and fourth quadrants) as well as negative (in the first and third quadrants), but this pattern collapses to the constant global value $\alpha(R_t, V_{t-1}^*) = \alpha_{\text{glob}} \approx 0$ as the bandwidths tend to infinity. In Figures 2c and 2d we can explore the conditional dependence between R_t and V_{t-1} at high and low levels of R_{t-1} , respectively, and we see even more clear differences in this dimension, especially in the first and second quadrants.

The p -values for tests of the hypothesis (J.2) and its opposite counterpart, using our new test for conditional independence as presented in the next section and with c -values for the bandwidth as chosen there, are both equal to 0.

K POWER STUDIES

Table 2* below is the extended version of Table 2 in the main article, sample size is $n = 100$:

↓ Test DGP →	Level				Power					
	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, $c=1$	0.096	0.060	0.048	0.072	0.668	0.756	0.388	0.860	0.828	0.680
HEL, $c=1.5$	0.068	0.056	0.052	0.056	0.888	0.940	0.512	0.924	0.952	0.812
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, $c=1$	0.046	0.050	0.050	0.047	0.746	0.717	0.400	0.873	0.566	0.320
MCC, $c=1.5$	0.040	0.052	0.056	0.055	0.814	0.779	0.329	0.889	0.618	0.341
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, $n = 100$. This is the full version of Table 2 in the main article

In Table 3* we present level and power results for various conditional independence tests applied to the DGPs listed in Table 1 in the paper for sample size $n = 200$:

↓ Test DGP →	Level				Power					
	1	2	3	4	5	6	7	8	9	10
BRT, $c=1$	0.037	0.025	0.028	0.032	0.995	0.996	<u>0.979</u>	0.989	0.904	0.785
BRT, $c=1.5$	0.044	0.025	0.025	0.037	0.971	0.993	0.931	0.997	0.931	0.759
BRT, $c=2$	0.064	0.023	0.023	0.052	0.943	0.979	0.873	0.997	0.912	0.728
BT, $c_1=1, c_2=1$	0.047	0.051	0.041	0.053	0.996	0.812	0.852	<u>1.000</u>	0.936	-
BT, $c_1=0.85, c_2=0.7$	0.048	0.044	0.064	0.056	0.988	0.728	0.792	<u>1.000</u>	0.908	-
BT, $c_1=0.75, c_2=0.6$	0.036	0.048	0.052	0.052	0.976	0.719	0.808	<u>1.000</u>	0.896	-
CHF	0.046	0.042	-	-	0.976	0.988	0.820	0.952	0.944	0.864
CM	0.044	0.056	0.060	0.048	0.992	0.740	0.788	0.680	0.476	0.360
HEL, $c=1$	0.064	0.052	0.080	0.080	0.900	0.960	0.596	0.992	0.968	0.880
HEL, $c=1.5$	0.064	0.056	0.048	0.036	0.980	<u>1.000</u>	0.808	0.992	0.972	0.972
HEL, $c=2$	0.044	0.060	0.056	0.048	<u>1.000</u>	<u>1.000</u>	0.864	<u>1.000</u>	<u>1.000</u>	<u>0.996</u>
KS	0.068	0.053	0.048	0.084	0.952	0.552	0.660	0.532	0.336	0.284
LGPC, $c=1.0$	0.039	0.052	0.054	0.054	0.995	0.948	0.818	<u>1.000</u>	<u>1.000</u>	0.985
LGPC, $c=1.4$	0.042	0.057	0.058	0.042	<u>1.000</u>	0.993	0.956	<u>1.000</u>	<u>1.000</u>	0.958
LIN	0.043	0.053	0.042	0.050	<u>1.000</u>	0.354	0.250	0.113	0.172	0.143
MCC, $c=1$	0.049	0.051	0.057	0.054	0.982	0.983	0.831	<u>1.000</u>	0.947	0.679
MCC, $c=1.5$	0.046	0.048	0.049	0.053	0.995	0.989	0.872	<u>1.000</u>	0.968	0.738
MCC, $c=2$	0.045	0.045	0.047	0.057	0.997	0.995	0.735	<u>1.000</u>	0.971	0.745
SCM	0.048	0.060	0.064	0.068	0.980	0.648	0.620	0.720	0.352	0.280
SEL	0.052	0.033	-	-	0.992	<u>1.000</u>	0.972	<u>1.000</u>	0.884	0.864
SKS	0.056	0.028	0.064	0.072	0.964	0.324	0.512	0.552	0.148	0.136

Table 3*: Level and power of conditional independence tests, $n = 200$

Su and White (2014) define two extensions to a subset of the data generating processes defined in Table 1 in which the conditioning variable \mathbf{X}_3 is a vector with two and three components respectively. The first extension turns the conditioning variable $\mathbf{X}_{3,t}$ in DGP1-DGP2 and DGP5-DGP10 $X_{3,t}$ into a bivariate vector, where we define DGP1' in the same way as before,

$$1'. (X_{1,t}, X_{2,t}, \mathbf{X}_{3,t}) = (\epsilon_{1,t}, \epsilon_{2,t}, \boldsymbol{\epsilon}_{3,t}),$$

but where $\boldsymbol{\epsilon}_{3,t} \sim \mathcal{N}(0, \mathbf{I}_2)$, and where we define DGP2' and DGP5'-DGP10' by setting $\mathbf{X}_{3,t} = (X_{1,t-1}, X_{1,t-2})$, keeping $X_{2,t}$ as before, and,

$$2'. X_{1,t} = 0.5X_{1,t-1} + 0.25X_{1,t-2} + \epsilon_{1,t},$$

$$5'. X_{1,t} = 0.5X_{1,t-1} + 0.25X_{1,t-2} + 0.5X_{2,t} + \epsilon_{1,t},$$

$$6'. X_{1,t} = 0.5X_{1,t-1} + 0.25X_{1,t-2} + 0.5X_{2,t}^2 + \epsilon_{1,t},$$

- 7'. $X_{1,t} = 0.5X_{1,t-1}X_{2,t} + 0.25X_{1,t-2} + \epsilon_{1,t}$,
- 8'. $X_{1,t} = 0.5X_{1,t-1} + 0.25X_{1,t-2} + 0.5X_{2,t}\epsilon_{1,t}$,
- 9'. $X_{1,t} = \sqrt{h_t}\epsilon_{1,t}$, $h_t = 0.01 + 0.5X_{1,t-1}^2 + 0.25X_{1,t-2}^2 + 0.25X_{2,t}^2$, and
- 10'. Same as DGP10 above, except for the new definition of $\mathbf{X}_{3,t}$.

We report the level and power results obtained by Su and White (2014) for the methods CHF, CM, KS and SEL on these data by testing the null hypothesis

$$H_0 : X_{t,1} \perp X_{t,2} \mid \mathbf{X}_{t,3},$$

and include results from our new test based on the multivariate simplification of the LGPC as defined in Section 3.1.2 in table 4*. The results are quite acceptable and compares well with other nonparametric methods.

The second extension introduced by Su and White (2014) increases the dimension of the conditioning variable $\mathbf{X}_{3,t}$ once more, so that $\epsilon_{3,t} \sim \mathcal{N}(0, \mathbf{I}_3)$, and where DGP2'' and DGP5''-DGP10'' are defined by by setting $\mathbf{X}_{3,t} = (X_{1,t-1}, X_{1,t-2}, X_{1,t-3})$, and,

- 2''. $X_{1,t} = 0.5X_{1,t-1} + 0.25X_{1,t-2} + 0.125X_{1,t-3} + \epsilon_{1,t}$,
- 5''. $X_{1,t} = 0.5X_{1,t-1} + 0.25X_{1,t-2} + 0.125X_{1,t-3} + 0.5X_{2,t} + \epsilon_{1,t}$,
- 6''. $X_{1,t} = 0.5X_{1,t-1} + 0.25X_{1,t-2} + 0.125X_{1,t-3} + 0.5X_{2,t}^2 + \epsilon_{1,t}$,
- 7''. $X_{1,t} = 0.5X_{1,t-1}X_{2,t} + 0.25X_{1,t-2} + 0.125X_{1,t-3} + \epsilon_{1,t}$,
- 8''. $X_{1,t} = 0.5X_{1,t-1} + 0.25X_{1,t-2} + 0.125X_{1,t-3} + 0.5X_{2,t}\epsilon_{1,t}$,
- 9''. $X_{1,t} = \sqrt{h_t}\epsilon_{1,t}$, $h_t = 0.01 + 0.5X_{1,t-1}^2 + 0.25X_{1,t-2}^2 + 0.125X_{1,t-3}^2 + 0.25X_{2,t}^2$, and
- 10''. Same as DGP10 and DGP10' above, except for the new definition of $\mathbf{X}_{3,t}$.

Again, we observe simulation results in Table 5*. Contrary to Su and White (2014), we have also run our test for $n = 100$ which reveals that we can obtain some power also in that case. All in all, the simplified test based on the LGPC performs mostly on par with other non-parametric tests in this setting as well.

Sample size	\downarrow Test DGP \rightarrow	Level		Power					
		1'	2'	5'	6'	7'	8'	9'	10'
$n = 100$	CHF	0.028	0.042	0.720	0.704	0.412	0.564	<u>0.460</u>	<u>0.556</u>
	CM	0.028	0.016	0.656	0.360	0.108	0.512	0.164	0.208
	KS	0.040	0.020	0.400	0.284	0.056	0.380	0.124	0.176
	LGPC, $c=1.75$	0.048	0.033	<u>0.951</u>	0.668	0.294	0.517	0.406	0.538
	SEL	0.052	0.040	0.844	<u>0.828</u>	<u>0.620</u>	<u>0.568</u>	0.440	0.528
$n = 200$	CHF	0.030	0.040	0.948	0.944	0.748	0.828	<u>0.724</u>	<u>0.860</u>
	CM	0.050	0.032	0.940	0.588	0.304	0.792	0.304	0.364
	KS	0.046	0.024	0.776	0.432	0.168	0.696	0.216	0.284
	LGPC, $c=1.75$	0.059	0.035	<u>0.998</u>	0.923	0.415	0.769	0.698	0.833
	SEL	0.058	0.026	0.972	<u>0.988</u>	<u>0.932</u>	<u>0.832</u>	0.684	0.832
$n = 400$	CHF	0.040	0.036	<u>1.000</u>	0.984	0.972	0.996	0.920	<u>0.984</u>
	CM	0.056	0.024	<u>1.000</u>	0.884	0.552	0.980	0.556	0.668
	KS	0.060	0.024	<u>1.000</u>	0.732	0.324	0.952	0.384	0.524
	LGPC, $c=1.75$	0.048	0.018	<u>1.000</u>	0.996	0.584	0.960	<u>0.928</u>	0.978
	SEL	0.040	0.030	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	0.836	0.884

Table 4*: Level and power, 4-dimensional data

Sample size	\downarrow Test DGP \rightarrow	Level		Power					
		1''	2''	5''	6''	7''	8''	9''	10''
$n = 100$	LGPC, $c=1.75$	0.068	0.049	<u>0.911</u>	<u>0.567</u>	<u>0.295</u>	<u>0.266</u>	<u>0.380</u>	<u>0.521</u>
$n = 200$	CHF	0.028	0.022	0.964	0.952	0.668	<u>0.852</u>	<u>0.552</u>	<u>0.856</u>
	CM	0.050	0.028	0.756	0.484	0.192	0.788	0.260	0.448
	KS	0.048	0.032	0.500	0.380	0.096	0.660	0.212	0.372
	LGPC, $c=1.75$	0.062	0.050	0.993	0.776	0.388	0.363	0.545	0.794
	SEL	0.056	0.026	<u>0.996</u>	<u>0.980</u>	<u>0.860</u>	0.816	0.344	0.680
$n = 400$	CHF	0.032	0.034	<u>1.000</u>	0.972	0.928	0.884	<u>0.792</u>	<u>0.972</u>
	CM	0.050	0.032	0.992	0.728	0.400	<u>0.912</u>	0.390	0.620
	KS	0.044	0.036	0.840	0.552	0.220	0.880	0.306	0.568
	LGPC, $c=1.75$	0.048	0.030	<u>1.000</u>	0.938	0.519	0.538	0.740	0.956
	SEL	0.056	0.038	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	0.888	0.616	0.876

Table 5*: Level and power, 5-dimensional data

REFERENCES

- Billingsley, Patrick. 2008. *Probability and Measure*. John Wiley & Sons.
- Brockwell, Peter J, and Richard A Davis. 2006. *Time Series: Theory and Methods: Theory and Methods*. Springer.
- 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.
- Fan, Jianqing, and Qiwei Yao. 2008. *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer Science & Business Media.
- Gonçalves, Silvia, and Halbert White. 2002. “The Bootstrap of the Mean for Dependent Heterogeneous Arrays.” *Econometric Theory* 18 (6): 1367–84.
- . 2004. “Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models.” *Journal of Econometrics* 119 (1): 199–219.
- Hall, P, and CC Heyde. 1980. “Martingale Limit Theory and Its Applications, Acad.” *Press, New York*.
- 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.
- Joe, Harry. 1989. “Estimation of Entropy and Other Functionals of a Multivariate Density.” *Annals of the Institute of Statistical Mathematics* 41 (4): 683–97.
- 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.
- . 2019. “Estimating and Testing Nonlinear Local Dependence Between Two Time Series.” *Journal of Business & Economic Statistics* 37 (4): 648–60.
- 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.
- Paparoditis, Efsthios, and Dimitris N Politis. 2000. “The Local Bootstrap for Kernel Estimators Under General Dependence Conditions.” *Annals of the Institute of Statistical Mathematics* 52 (1): 139–59.
- Rio, Emmanuel. 1995. “The Functional Law of the Iterated Logarithm for Stationary Strongly Mixing Sequences.” *The Annals of Probability*, 1188–1203.
- Schervish, Mark J. 2012. *Theory of Statistics*. Springer Science & Business Media.
- Su, Liangjun, and Halbert White. 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.
- 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.
- Van den Bos, Adriaan. 2007. *Parameter Estimation for Scientists and Engineers*. John Wiley & Sons.
- Wang, Xia, and Yongmiao Hong. 2018. "Characteristic Function Based Testing for Conditional Independence: A Nonparametric Regression Approach." *Econometric Theory* 34 (4): 815–49.
- Yahoo Finance. 2018. *Accessed June 2018*.