Some notes for the revision of the JBES paper 2020

Version: April 21 2020

Notation: $Z = [Z_1, Z_2, Z_3]$

Z_1 is a d_1-vector; Z_2 is a d_2-vector; Z_3 is a d_3-vector.

Local means: $\mu_1(z_1)$ is d_1-vector; $\mu_2(z_2)$ is d_2-vector; $\mu_3(z_3)$ is a d_3-vector.

Local covariance matrices (assuming pairwise z-coordinate dependence):

$$\mathbf{\Sigma}_{ij} = \mathrm{cov}(Z_i, Z_j) = (\sigma_{kl}(z_i^k, z_j^l)) \quad \text{for} \quad i,j = 1,2,3 \;\; k = 1, \dots, d_i; l = 1, \dots, d_j$$

this is a $d_i \times d_j$ matrix. Then, omitting in our notation the dependence on the local coordinate z,

$$\mathrm{cov}(Z) = \mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} & \mathbf{\Sigma}_{13} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} & \mathbf{\Sigma}_{23} \\ \mathbf{\Sigma}_{31} & \mathbf{\Sigma}_{32} & \mathbf{\Sigma}_{33} \end{bmatrix}$$

is a matrix of dimension $(d_1 + d_2 + d_3) \times (d_1 + d_2 + d_3)$, such that $\mathbf{\Sigma}^T = \mathbf{\Sigma}$, $\mathbf{\Sigma}_{12}^T = \mathbf{\Sigma}_{21}$, $\mathbf{\Sigma}_{13}^T = \mathbf{\Sigma}_{31}$, $\mathbf{\Sigma}_{23}^T = \mathbf{\Sigma}_{32}$. The local Gaussian can now be written (again omitting z in parameters),

$$f(z) = \frac{1}{(2\pi)^{d_1+d_2+d_3)/2}} \exp[-\frac{1}{2}(z - \mu)^T \mathbf{\Sigma}^{-1}(z - \mu)] \tag{1}$$

Moreover,

$$f(z_1, z_2|z_3) = \frac{1}{(2\pi)^{d_1+d_2)/2}} \exp[-\frac{1}{2}[(z_{12|3} - \mu_{12|3})^T \mathbf{\Sigma}_{12|3}^{-1}(z_{12|3} - \mu_{12|3})]$$

where $z_{12|3}^T = [z_1 z_2]$. To define $\mu_{12|3}$ and $\mathbf{\Sigma}_{12|3}$, let $\mathbf{\Sigma}^{11}$, $\mathbf{\Sigma}^{12}$, $\mathbf{\Sigma}^{21}$ and $\mathbf{\Sigma}^{22}$ 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

$$\mathbf{\Sigma}^{11} = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix}, \quad \mathbf{\Sigma}^{12} = \begin{bmatrix} \mathbf{\Sigma}_{13} \\ \mathbf{\Sigma}_{23} \end{bmatrix}, \quad \mathbf{\Sigma}^{21} = \begin{bmatrix} \mathbf{\Sigma}_{31} & \mathbf{\Sigma}_{32} \end{bmatrix}, \quad \mathbf{\Sigma}^{22} = \mathbf{\Sigma}_{33}$$

such that

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}^{11} & \mathbf{\Sigma}^{12} \\ \mathbf{\Sigma}^{21} & \mathbf{\Sigma}^{22} \end{bmatrix}$$

Then

$$\mu_{12|3} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \mathbf{\Sigma}^{12}(\mathbf{\Sigma}^{22})^{-1}(z_3 - \mu_3) = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \mathbf{\Sigma}_{13}\mathbf{\Sigma}_{33}^{-1}(z_3 - \mu_3) \\ \mathbf{\Sigma}_{23}\mathbf{\Sigma}_{33}^{-1}(z_3 - \mu_3) \end{bmatrix} \tag{2}$$

Further,

$$\mathbf{\Sigma}_{12|3} = \mathbf{\Sigma}^{11} - \mathbf{\Sigma}^{12}(\mathbf{\Sigma}^{22})^{-1}\mathbf{\Sigma}^{21} = \mathbf{\Sigma}^{11} - \begin{bmatrix}\mathbf{\Sigma}_{13}\\ \mathbf{\Sigma}_{23}\end{bmatrix}\mathbf{\Sigma}_{33}^{-1}\begin{bmatrix}\mathbf{\Sigma}_{31} & \mathbf{\Sigma}_{32}\end{bmatrix}$$

$$= \begin{bmatrix}\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{13}\mathbf{\Sigma}_{33}^{-1}\mathbf{\Sigma}_{31} & \mathbf{\Sigma}_{12} - \mathbf{\Sigma}_{13}\mathbf{\Sigma}_{33}^{-1}\mathbf{\Sigma}_{32}\\ \mathbf{\Sigma}_{21} - \mathbf{\Sigma}_{23}\mathbf{\Sigma}_{33}^{-1}\mathbf{\Sigma}_{31} & \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{23}\mathbf{\Sigma}_{33}^{-1}\mathbf{\Sigma}_{32}\end{bmatrix} \tag{3}$$

Next, to set up conditional independence,

$$f(z_1|z_3) = \frac{1}{(2\pi)^{d_1/2}|\mathbf{\Sigma}_{1|3}|^{1/2}}\exp\{-\frac{1}{2}(z_1-\mu_{1|3})^T\mathbf{\Sigma}_{1|3}^{-1}(z_1-\mu_{1|3})\}$$

$$f(z_2|z_3) = \frac{1}{(2\pi)^{d_2/2}|\mathbf{\Sigma}_{2|3}|^{1/2}}\exp\{-\frac{1}{2}(z_2-\mu_{2|3})^T\mathbf{\Sigma}_{2|3}^{-1}(z_2-\mu_{2|3})\}$$

with

$$\mu_{1|3} = \mu_1 + \mathbf{\Sigma}_{13}\mathbf{\Sigma}_{33}^{-1}(z_3-\mu_3) \quad \mu_{2|3} = \mu_2 + \mathbf{\Sigma}_{23}\mathbf{\Sigma}_{33}^{-1}(z_3-\mu_3) \tag{4}$$

and

$$\mathbf{\Sigma}_{1|3} = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{13}\mathbf{\Sigma}_{33}^{-1}\mathbf{\Sigma}_{31} \quad \mathbf{\Sigma}_{2|3} = \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{23}\mathbf{\Sigma}_{33}^{-1}\mathbf{\Sigma}_{32} \tag{5}$$

Under conditional independence of Z_1 and Z_2 given Z_3,

$$f(z_1, z_2|z_3) = f(z_1|z_3)f(z_2|z_3)$$

$$= \frac{1}{(2\pi)^{(d_1+d_2)/2}|\mathbf{\Sigma}_{1|3}|^{1/2}|\mathbf{\Sigma}_{2|3}|^{1/2}}\exp\{-\frac{1}{2}[(z_1-\mu_{1|3})^T\mathbf{\Sigma}_{1|3}^{-1}(z_1-\mu_{1|3})+(z_2-\mu_{2|3})^T\mathbf{\Sigma}_{2|3}^{-1}(z_2-\mu_{2|3})]\}$$

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

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

Let $\mathbf{D} = \mathrm{diag}\{\mathbf{\Sigma}_{12|3}\}$. Then the partial correlation matrix may be defined as the lower left hand block matrix of $\mathbf{D}^{-1/2}\mathbf{\Sigma}_{12|3}\mathbf{D}^{-1/2}$. CHECK

Some special cases:

a) $d_1 = d_2 = d_3 = 1$:

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

Here, since the Z_i-s are normalized with variance one,

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

and

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

Hence, using symmetry

$$\mathbf{\Sigma}^{11} - \mathbf{\Sigma}^{12}(\mathbf{\Sigma}^{22})^{-1}\mathbf{\Sigma}^{21} = \begin{bmatrix} 1 & \rho_{12} \\ \rho_{21} & 1 \end{bmatrix} - \begin{bmatrix} \rho_{13} \\ \rho_{23} \end{bmatrix} \begin{bmatrix} \rho_{31} & \rho_{32} \end{bmatrix}$$

$$= \begin{bmatrix} 1-\rho_{13}^2 & \rho_{12}-\rho_{13}\rho_{23} \\ \rho_{12}-\rho_{13}\rho_{23} & 1-\rho_{23}^2 \end{bmatrix}$$

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

$$\begin{bmatrix} 1 & \frac{\rho_{12}-\rho_{13}\rho_{23}}{\sqrt{1-\rho_{13}^2}\sqrt{1-\rho_{23}^2}} \\ \frac{\rho_{12}-\rho_{13}\rho_{23}}{\sqrt{1-\rho_{13}^2}\sqrt{1-\rho_{23}^2}} & 1 \end{bmatrix} \qquad (*)$$

Alternatively this can be obtained by

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

Another special case: $d_3 = 1$, d_1 and d_2 arbitrary. We have

$$\mathbf{\Sigma}^{11} = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix}$$

where

$$\mathbf{\Sigma}_{11} = \{\rho_{ij}^{1,1}\},\ i = 1,\ldots,d_1; j = 1,\ldots,d_1 \quad \mathbf{\Sigma}_{12} = \{\rho_{ij}^{1,2}\},\ i = 1,\ldots,d_1; j = 1,\ldots,d_2$$

$$\mathbf{\Sigma}_{21} = \{\rho_{ij}^{2,1}\},\ i = 1,\ldots,d_2; j = 1,\ldots,d_1 \quad \mathbf{\Sigma}_{22} = \{\rho_{ij}^{2,2}\},\ ,i = 1,\ldots,d_2; j = 1,\ldots,d_2$$

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

$$\mathbf{\Sigma}^{12}(\mathbf{\Sigma}^{22})^{-1}\mathbf{\Sigma}^{21} = \mathbf{\Sigma}^{12}\mathbf{\Sigma}^{21} = \begin{bmatrix} \mathbf{\Sigma}_{13} \\ \mathbf{\Sigma}_{23} \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}_{31} & \mathbf{\Sigma}_{32} \end{bmatrix}$$

where

$$\mathbf{\Sigma}_{13} = \{\rho_{i3}^{1,3}\},\ i = 1,\ldots,d_1 \quad \mathbf{\Sigma}_{23} = \{\rho_{j3}^{2,3}\},\ j = 1,\ldots,d_2$$

$$\mathbf{\Sigma}_{31} = \{\rho_{3i}^{3,1}\},\ i = 1,\ldots,d_1 \quad \mathbf{\Sigma}_{32} = \{\rho_{3j}^{3,2}\},\ j = 1,\ldots,d_2$$

Further,

$$\mathbf{\Sigma}_{12|3} = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix} - \begin{bmatrix} \mathbf{\Sigma}_{13} \\ \mathbf{\Sigma}_{23} \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}_{31} & \mathbf{\Sigma}_{32} \end{bmatrix}$$

Along the diagonal of $\mathbf{\Sigma}_{11}$ and $\mathbf{\Sigma}_{22}$ there are ones along the diagonal, whereas $\begin{bmatrix} \mathbf{\Sigma}_{13} \\ \mathbf{\Sigma}_{23} \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}_{31} & \mathbf{\Sigma}_{32} \end{bmatrix}$ has $(\rho_{i3}^{1,3})^2,\ i = 1,\ldots,d_1$ and $(\rho_{3j}^{3,2})^2,\ j = 1,\ldots,d_2$ along the diagonal such that the $\mathbf{D}$ is the $(d_1+d_2)\times(d_1+d_2$ diagonal matrix with the concatenated sequence $1-(\rho_{i3}^{1,3})^2,\ i=1,\ldots,d_1; 1-(\rho_{3j}^{3,2})^2,\ j=1,\ldots,d_2$ along the diagonal.

Pre -and postmultiplying in $\mathbf{\Sigma}_{12|3}$ by $\mathbf{D^{-1/2}}$ converts $\mathbf{\Sigma}_{12|3}$ into a correlation matrix. All of elements in the off-diagonal elements of this matrix has the form of the off-diagonal elements of (*), such that

$$\rho(Z_{1,i},Z_{1,j}|Z_3)=\frac{\rho(Z_{1,i},Z_{1,j})-\rho(Z_{1,i},Z_3)\rho(Z_{1,j},Z_3)}{\sqrt{1-\rho^2(Z_{1,i},Z_3)}\sqrt{1-\rho^2(Z_{1,j},Z_3)}},\ i,j=1,\ldots,d_1 \quad (**)$$

$$\rho(Z_{2,i},Z_{2,j}|Z_3)=\frac{\rho(Z_{2,i},Z_{2,j})-\rho(Z_{2,i},Z_3)\rho(Z_{2,j},Z_3)}{\sqrt{1-\rho^2(Z_{2,i},Z_3)}\sqrt{1-\rho^2(Z_{2,j},Z_3)}},\ i,j=1,\ldots,d_2; \quad (***)$$

$$\rho(Z_{1,i},Z_{2,j}|Z_3)=\frac{\rho(Z_{1,i},Z_{2,j})-\rho(Z_{1,i},Z_3)\rho(Z_{2,j},Z_3)}{\sqrt{1-\rho^2(Z_{1,i},Z_3)}\sqrt{1-\rho^2(Z_{2,j},Z_3)}},\ i=1,\ldots,d_1; j=1,\ldots,d_2 \quad (****)$$

Testing for conditional independence between Z_1 and Z_2 given Z_3:

It seems to be enough to test for zero correlations for all pairs of correlations $\rho(Z_{1,i}(z_1),Z_{2,j}(z_2)|Z_3(z_3))$, and we have already tests for doing that. One can also, based on (ii) and (iii), test for for conditional independence between other combinations of components, given Z_3. Note that (**), (***), and (****) give explicit expressions for the conditional correlation matrix of Z_1 and Z_2 given Z_3.

Another special case: $d_1=d_2=1$, d_3 arbitrary:

In this case,

$$\mathbf{\Sigma}^{11}=\begin{bmatrix}1 & \rho_{12}\\ \rho_{21} & 1\end{bmatrix} \quad \mathbf{\Sigma}_{13}=\{\rho_{1i}^{1,3}\}, i=1,\ldots,d_3; \quad \mathbf{\Sigma}_{23}=\{\rho_{2,i}^{2,3}\},\ i=1,\ldots,d_3.$$

hence $\begin{bmatrix}\mathbf{\Sigma}_{13}\\ \mathbf{\Sigma}_{23}\end{bmatrix}$ is of dimension $2\times d_3$, whereas $\begin{bmatrix}\mathbf{\Sigma}_{31} & \mathbf{\Sigma}_{32}\end{bmatrix}$ is of dimension $d_3\times 2$ such that

$$\begin{bmatrix}\mathbf{\Sigma}_{13}\\ \mathbf{\Sigma}_{23}\end{bmatrix}\mathbf{\Sigma}_{33}^{-1}\begin{bmatrix}\mathbf{\Sigma}_{31} & \mathbf{\Sigma}_{32}\end{bmatrix}$$

is of dimension 2×2. But it is immediately seen that no formulas like (**), (***) and (****) will appear. In fact it will be much more complicated with sums and products of ρ-s. This is consistent with the partial correlation entry in Wikipedia where they propose using recursive formulas for obtaining partial correlations in the general vector case.

General case: d_1,d_2,d_3 arbitrary: Seems impossible to obtain explicit formulas, but of course the general fact that the local partial correlation matrix of Z_1 and Z_2 given Z_3, can be identified by the lower left hand block matrix of $\mathbf{D}^{-1/2}\mathbf{\Sigma}_{12|3}\mathbf{D}^{-1/2}$ remains valid.

Estimation

In the general case the estimate of the partial correlation matrix is given by the lower left hand block matrix of (cf. page 2)

$$\hat{\mathbf{D}}^{-1/2}\hat{\mathbf{\Sigma}}_{12|3}\hat{\mathbf{D}}^{-1/2} = \hat{\mathbf{D}}^{-1/2}\begin{bmatrix}\hat{\mathbf{\Sigma}}_{11} - \hat{\mathbf{\Sigma}}_{13}\hat{\mathbf{\Sigma}}_{33}^{-1}\hat{\mathbf{\Sigma}}_{31} & \hat{\mathbf{\Sigma}}_{12} - \hat{\mathbf{\Sigma}}_{13}\hat{\mathbf{\Sigma}}_{33}^{-1}\hat{\mathbf{\Sigma}}_{32} \\ \hat{\mathbf{\Sigma}}_{21} - \hat{\mathbf{\Sigma}}_{23}\hat{\mathbf{\Sigma}}_{33}^{-1}\hat{\mathbf{\Sigma}}_{31} & \hat{\mathbf{\Sigma}}_{22} - \hat{\mathbf{\Sigma}}_{23}\hat{\mathbf{\Sigma}}_{33}^{-1}\hat{\mathbf{\Sigma}}_{32}\end{bmatrix}\hat{\mathbf{D}}^{-1/2}$$

let us block divide the $(d_1 + d_2) \times (d_1 + d_2)$ diagonal matrix $\mathbf{D}$ into

$$\mathbf{D} = \begin{bmatrix}\mathbf{D}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{22}\end{bmatrix}$$

then the estimated partial correlation matrix of Z_1 and Z_2 given Z_3 is given by

$$\hat{\mathbf{D}}_{11}^{-1/2}(\hat{\mathbf{\Sigma}}_{21} - \hat{\mathbf{\Sigma}}_{23}\hat{\mathbf{\Sigma}}_{33}^{-1}\hat{\mathbf{\Sigma}}_{31})\hat{\mathbf{D}}_{22}^{-1/2} \qquad (a)$$

As stated above, in case Z_3 is a vector, it is difficult to find an explicit expression of the partial correlation matrix in terms of pairwise correlations. If Z_3 is a scalar, then the expression (****) can be used where the partial correlation between $Z_{1,i}$ and $Z_{2,j}$ given Z_3 is expressed in terms of the three partial correlations $\{\rho(Z_{1i}, Z_{2j}, \rho(Z_{1i}, Z_3), \rho(Z_{2j}, Z_3)\}$ where $i = 1, \ldots, d_1$ and $j = 1, \ldots, d_2$.

Asymptotic estimation theory:

This has been treated in *OT19* for some situations:

(i) $d_1 = d_2 = d_3 = 1$, three coordinates $z = (z_1, z_2, z_3)$; this has been given in Theorem 3.1 in *OT19*.

(ii) $d_1 = d_2 = 1$, d_3 arbitrary but with pairwise variables and coordinates. The asymptotics is generally given in (see (20) *OT19)*

$$\sqrt{nb_n^2}(\hat{\alpha}(z)) - \alpha(z) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nabla g(\rho)^T \mathbf{\Omega} \nabla g(\rho)) \qquad (b)$$

Here $\mathbf{\Omega}$ is defined in Theorem 2 of *OT18*. The function g is the function taking the relevant vector of pairwise correlations into the partial correlation as for example in (8) of *OT19*. The computation of ∇g is outlined in the supplementary material of *OT19*.

The results can be extended to the new situations considered in this note. The assumptions in the theorem is the same as in Theorem 3.1 of *OT19*.

(iii) d_1 and d_2 arbitrary, $d_3 = 1$. We are then in the situation of (****). As in the above limit results, the limit distribution of $\hat{\rho}(Z_{1i}, Z_{2j}|Z_3)$ follows from the delta rule applied to the joint distribution of $\{\hat{\rho}(Z_{1i}, Z_{2j}), \hat{\rho}(Z_{1i}, Z_3), \hat{\rho}(Z_{2j}, Z_3)$. The joint distribution for these three correlation estimates follow again directly from Theorem 2 in *OT19*, where it is noted that the covariance between two correlation estimates is of lower order if the two index pairs are distinct. The variance from the limiting normal distribution can then be found by computing ∇g as in (b). In fact ∇g is easy to compute directly from taking appropriate derivatives in (****).

(iv) This is the general case where d_1, d_2, d_3 are all arbitrary. Then in principle for each $i = 1, \dots, d_1$ and $j = 1, \dots, d_2$, $\rho(Z_{1i}, Z_{2j}|Z_3$, using equation (a), can be expressed as a deterministic function g_{ij} of a correlation vector ρ and asymptotic normality can be proved as before using the delta rule and Theorem 2 of *OT19*. It is difficult however to find an explicit expression for the variance in that distribution because the complexity of (a) in the matrix differentiations that have to be carried out to determine ∇g_{ij}.

The joint asymptotic normality of the pairwise partial correlation estimates can be proved by use of the Cramer-Wold device. But, again, it is hard to obtain an explicit expression for the covariance matrix of that distribution is the general case.

Finally, it should be observed that a more general pairwise scheme can be devised. This has been used in *LT18* and *LT19* and in *JT19*.

(v) This is the scheme were the pairwise set-up is used not only for the local correlations (1 parameter) but for the 5 parameters of local correlation $\rho(x_1, x_2)$, local variances $\sigma_i(x_1, x_2), i = 1, 2,$ $\mu_i(x_1, x_2), i = 1, 2$. This really amounts to a pairwise normal approximation to the multivariate distribution. For more than two variables, possibly the variances and means should be replaced by $\sigma_i(x_i)$ and $\mu_i(x_i)$. The asymptotic theory is given in *TH13*. This has been written on x-level, but may also be made valid on the z-level. Have to be a bit careful here in the derivations for the bootstrap to follow next.

The validity of the bootstrap: estimation

The asymptotic theory for the local partial correlation involves complicated expressions which may be difficult to implement in practice. It is natural to try the bootstrap instead to construct confidence intervals. Since we allow a situation where the $Z = \{Z_t\}$ is a time series one has in general to use the block bootstrap, and one is faced with the problem of showing the validity of the bootstrap.

The bootstrap validity problem was studied on the x-level 5-parameter case, as in (v) above, for two variables (X_t, Y_{t-k}) for a bivariate stationary time series $\{X_t, Y_t\}$, and it was indicated that the results holds true also on the simpler z-level. (The results of Theorem 3.2 in *OT19* are crucial here.)

It has been seen that the building blocks of the asymptotic theory of the local partial correlation $\rho(Z_{1i}, Z_{2j})|Z_3)$ are the local correlations $\rho(Z_{ki}, Z_{lj})$

where $k, l = 1, 2, 3$ and where i and j range in the interval $1, \ldots, d_1; 1, \ldots, d_2; 1, \ldots, d_3$. Let us simply use the notation ρ_{ij} for one of these local correlations, and let $\hat{\rho}_{ij}$ for the pairwise local likelihood estimate as outlined in *OT17* (or *LT18*). If we can prove that the block bootstrap is valid for such a pairwise local likelihood estimate, then it follows, relatively straighforwardly, that it is also valid for the appropriate vector of such pairwise estimates used in the above asymptotic theory in the cases (i) - (iv). (The situation in (v) need some more care for the arguments of σ_i and μ_i). Next the validity of the local partial correlation can be obtained as a deterministic function of the elements of this vector of local correlations.

We can use the bootstrap theory of *LT19*, in itself based on Gonçcalves and White (2002, 2004), more or less directly, since for a given local correlation estimate $\hat{\rho}_{ij}$ the situation treated in Gonçalves and White is very similar to *LT19*. We spell out some details to be more explicit. First we need to specify the analogue of the conditions A1-A5 in *LT19*, where the parameter θ now is replaced by ρ_{ij}, and the time series $\{X_t, Y_t\}$ by $\{Z_t\}$ with $Z_t = [Z_{1t}, Z_{2t}, Z_{3t}]$. (In case this vector time series consists of iid observations a simplified theory as indicated in *LT18* can be used).

We need to introduce some more notation to formulate the conditions: To ease the comparison with *LT19* we use the notation (X_t, Y_t) ***(Seen in retrospect I am not so sure this is so smart. WE should probably have some uniform Z-notation)*** for the particular pair of time series consisting of components of $\{Z_t\}$ with a parameter given by $\theta = \rho_{ij}$. Note also that we can use the normalized time series $\{Z_t\}$ instead of $\{\hat{Z}_t\}$ because the difference between these is a smaller order effect as seen in Theorem 3.2 of *OT19*. The pairwise local log likelihood is given by

$$L_{n,b}((\underline{X}_n, \underline{Y}_n), \theta_b(x,y)) = \frac{1}{n} \sum_{t=1}^{n} K_b(X_t - x, Y_t - y) \log(\psi((X_t, Y_t), \theta_b(x,y)))$$

$$- \int K_b(v - x, w - y) \psi((v,w), \theta_b(x,y)) dv dw \qquad (c)$$

Here $\underline{X}_n = (X_1, \ldots, X_n)$, $\underline{Y} = (Y_1, \ldots, Y_n)$, n is the number of observations, $x, y, v, w \in \mathbb{R}$, $\theta_b(x,y) = \rho_{ij,b}$ is the local correlation parameter, $K_b(v-x, w-y) = (b_1 b_2)^{-1} K(\frac{v-x}{b_1}) K(\frac{w-y}{b_2})$ 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 $\int K(v,w) dv dw = 1$. For a fixed bandwidth b, $\theta_{0,b}$ is the true local parameter, that is (cf. Lacal and Tjøstheim (2017), the minimizer of the penalty function

$$\int K_b(v - x, w - y)[\psi((v,w), \theta(x,y)) - \log(\psi((v,w), \theta(x,y))) f(v,w)] dv dw \qquad (cc)$$

where f is the density function of (X_t, Y_t). Note in particular that $\theta_{0,b} \to \theta_0$ as $b \to 0$, which is the true local correlation parameter. The reader my consult Tjøstheim and Hufthammer (2013) for more details and regularity conditions. The corresponding estimate $\theta_{n,b}(x,y)$ is obtained by maximizing the log likelihood in (c). Note that

$$L_{n,b}(\theta) = \frac{1}{n}\sum_{t=1}^{n} L^{(b)}((X_t,Y_t),\theta) \tag{d}$$

where

$$L^{(b)}(X_t,Y_t,\theta) = K_b(X_t - x, Y_t - y)\log(\psi((X_t,Y_t),\theta)) - \int K_b(v-x,w-y)\psi((v,w),\theta)dvdw. \tag{e}$$

We also need a couple of new concepts before we are able to state the conditions A1-A5: For a given function g, the process $\{g(X_t,Y_t),\theta)\}$ is defined to be r-dominated on Θ uniformly in $t,n,t=1,\ldots,n$ if there exists a function $D_t : \mathbb{R} \to \mathbb{R}$ such that $|g((X_t,Y_t),\theta)| \leq D_t$, for all $\theta \in \Theta, t=1,\ldots,n$, D_t is $\mathcal{B}$-measurable ($\mathcal{B}$ the Borel σ-field) such that $||D_t||_r \doteq (\int |D_t|^r dt)^{1/r} \leq \Delta < \infty$, for all $t = 1,\ldots,n$. In the present case (x,y) is fixed in $\theta(x,y)$ and $\Theta = [-1,1]$. Moreover, the process $\{g(X_t,Y_t),\theta)\}$ is defined to be Lipshitz continuous on Θ, if

$$|g((X_t,Y_t),\theta) - g((X_t,Y_t),\theta')| \leq C_t|\theta - \theta'| \text{ a.s.} - \mathbb{P}$$

for all $\theta,\theta' \in \Theta$ and with a sufficiently large constant M such that $C_t \leq M$. Here $\mathbb{P}$ is the probability measure on the probability space on which $\{X_t,Y_t\}$ is defined.

We are now in a position to state the conditions A1 - A3 which are basic in proving consistency and asymptotic normality of $\hat{\theta} = \hat{\rho}_{ij}$. A slightly different set of conditions are stated in *OT18*, *OT19*, (it seems that consistency is not treated in *OT19*?

A1: $(\Omega,\mathcal{F},\mathbb{P})$ is a complete probability space on which is defined the sequence of random functions $\{L_{n,b} : \Omega \times \Theta \to \bar{\mathbb{R}}\}$. The parameter set Θ is compact (trivially fulfilled in our case since $\rho \in [0,1]$). The random process $\{X_t,Y_t\}$ is $\mathcal{F}_t$-measurable, stationary and α-mixing with $\alpha_k = O(\alpha^k), \alpha \in (0,1)$ for all t, and $\bar{\mathbb{R}}$ is the extended real line, and with bivariate density having support on all of $\mathbb{R}^2$.

A2: $\theta_{0,b}$ is the unique maximizer of $\mathbb{E}(L_{n,b}(\theta)) : \Theta \to \bar{\mathbb{R}}$ with $\mathbb{E}(L_n, b(\theta))$ as in (cc).

A3: For a fixed b, using stationarity,

$$J_{n,b} = \mathbb{E}(-\nabla^2 L_{n,b}(\theta_{0,b})) = J_b = \mathbb{E}(-\nabla^2 L^{(b)}((X_t,Y_t),\theta_{0,b}))$$

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

$$M_{n,b} = \text{var}(\sqrt{nb_1b_2}\nabla L_{n,b}(\theta_{0,b}))$$

is $O(1)$ and uniformly positive. Here ∇ is the gradient with respect to θ; in the present scalar parameter case, simply the differential operator $d/d\theta = d/d\rho_{ij}$. Due to stationarity $M_{n,b} = \mathrm{var}(\nabla L^{(b)}(X_t, Y_t), \theta_{0,b}))$. CHECK and compare to Theorem 3 in *OT18*.

The conditions A1 - A3 are quite mild and are discussed in *LT19*. Note that they have been formulated for the case where θ is one-dimensional and equal to ρ_{ij}. To have them fulfilled for all possible combinations i, j condition A1 for example can very easily be stated in terms of the $\{Z_t\}$ process. The way the conditions are stated it is trivial to extend them to a five parameter θ-case. See *LT19* for this. Since the consistency does not seem to have been explicitly stated in *OT19*, we state it here. It is needed in the consistency proof of the bootstrap.

Theorem (Consistency): Let assumption A1 and A2 hold, and let $\{L^{(b)}(X_t, Y_t), \theta)\}, \theta \in \Theta$, be Lipshitz continuous on Θ and 4-dominated on Θ uniformly in t, n. Moreover, assume that as $b = b_n \to 0$, $\sigma_b^2 \doteq \mathrm{var}(L^{(b)}(X_t, Y_t), \theta) \to \sigma^2 < \infty$ and $\mu_b \doteq \mathbb{E}(L^{(b)}(X_t, Y_t), \theta)) \to \mu < \infty$ for some constants σ^2 and μ, for all $t = 1, \ldots, n$ and $\theta \in \Theta$. Then $\theta_{n,b} - \theta_{0,b} \xrightarrow{\mathbb{P}} 0$, where θ_0 is the limiting value of $\theta_{0,b}$ as $b \to 0$.

The proof is as in the Appendix of *LT19* with trivial modifications to take into account the scalar nature of the parameter $\theta = \rho_{ij}$ in the present case.

To state the bootstrap results we need a few more concepts and two additional conditions. We block-bootstrap the entire vector series $\{Z_t\}$. This automatically leads to the same block bootstrap of any pair of components, which we have denoted by $\{X_t, Y_t\}$ in the above to ease the comparison with *LT19*. That paper has used both the stationary bootstrap and the ordinary block bootstrap. As in that paper, the results below hold irrespective of the choice of block bootstrap, and we will therefore not distinguish between them in the following, but refer to *LT19* for details on the bootstrap construction. Bootstrap quantities will be starred in the sequel.

The bootstrap is supposed to live on the probability space $(\Lambda, \mathcal{G}, \mathbb{P}^*)$. The outcomes $\lambda \in \Lambda$ depends of course on the outcomes $\omega \in \Omega$ in the original probability space. We are now ready to formulate the conditions A4 and A5:

A4: The bootstrap probability space $(\Lambda, \mathcal{G}, \mathbb{P}^*)$ is a complete probability space for all $\omega \in \Omega$, and $\{L^*_{n,b} : \Lambda \times \Omega \times \Theta \to \bar{\mathbb{R}}\}$ is a sequence of random functions such that $L^*_{n,b}(\theta) = L_{n,b}((\underline{X}^*_n(\lambda, \omega), \underline{Y}^*_n(\lambda, \omega)), \theta)$ where $X^*_t(\lambda, \omega) = X_{\tau^Z_t(\lambda)}(\omega)$, $Y^*_t(\lambda, \omega) = Y_{\tau^Z_t(\lambda)}(\omega)$, where $\tau^Z_t : \Lambda \to \mathbb{N}$, $\omega \in \Omega$, $\lambda \in \Lambda$ are vectors of random indices representing the block bootstrap operation on $\{Z_t\}$. Moreover, the block length l_n is such that $l_n = o(\sqrt{n})$ as $n \to \infty$.

A5: For $\theta \in \Theta$ and for every $t = 1, \ldots, n$, $\{L^{(b)}(X_t, Y_t), \theta\}$ is Lipschitz continuous on Θ, $\{\nabla L^{(b)}((X_t, Y_t), \theta)\}$ is 6-dominated on Θ uniformly in t, n, and $\{\nabla^2 L^{(b)}((X_t, Y_t), \theta)\}$ is Lipschitz continuous on Θ and 2-dominated on Θ uniformly in t, n.

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. Again we refer to *LT19* for more discussion.

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

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

Further, we write $Y^*_n \xrightarrow{d_{\mathbb{P}^*_\omega}} \mathcal{N}(0,1)$ prob $-\, \mathbb{P}$, if for every sub-sequence $\{n'\}$, there exits a further sub-sequence $\{n''\}$ such that $Y_{n''} \xrightarrow{d_{\mathbb{P}^*_\omega}} \mathcal{N}(0,1)$ a.s. (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 (2012)).

Next, we can state the validity of bootstrap in terms of consistency and a central limit theorem:

Theorem A: Let assumptions A1, A2 and A4 hold, and let $\{L^{(b)}((X_t, Y_t), \theta)\}$, $\theta \in \Theta$ be Lipschitz continuous 4-dominated on Θ uniformly in t, n. Moreover, assume that, as $b = b_n \to 0, \sigma^2_b \doteq \mathrm{var}(L^{(b)}((X_t, Y_t), \theta) \to \sigma^2 < \infty$ and $\mu_b \doteq \mathbb{E}(L^{(b)}(X_t, Y_t), \theta)) \to \mu < \infty$, for all $t = 1, \ldots, n$ and $\theta \in \Theta$. Then

$$\theta^*_{n,b} - \theta_{b,n} \overset{\mathbb{P}^*_\omega, \mathbb{P}}{\to} 0 \qquad (x)$$

.

Theorem B: Let the assumptions of (i) - (v) of Theorems 3.1 and 3.2 (in *OT19*) and A1-A5 hold. Then

$$\sqrt{nb_1b_2} M^{-1/2}_{n,b} J_{n,b}(\theta^*_{n,b} - \theta_{n,b}) \overset{d_{\mathbb{P}^*_\omega}}{\to} \mathcal{N}(0,1)\mathrm{prob} - \mathbb{P}. \qquad (y)$$

Note that the theorem is formulated in the simplest case here, where there is only one parameter. A more complicated version is given in *LT19* in the 5-parameter case. The proofs are all essentially the same and is spelled out in some detail in the supplementary material of *LT19* as Theorems D1 and D2.

Theorem A and B are concerned with just one local correlation. But the local partial correlation 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) on pages 5 and 6 of this note. In particular alternative (iii), expressed more fully in equation (****), with $d_3 = 1$ and d_1 and d_2 arbitrary, and alternative (i), where $d_1 = d_2 = d_3 = 1$, but with local correlations depending on three coordinates (z_1, z_2, z_3), will be used as illustrations in the sequel. But the Corollaries A and B below are for a general situation with a pairwise model and pairwise coordinates including the case of alternative (i)$'$ which is (i) but with retsriction to pairwise coordinates. Corollary A treating bootstrap consistency and Corollary B bootstrap CLT.

Corollary A: Assume the conditions of Theorem A. Let $\alpha_{n,b}(i,j|Z_3) \doteq \hat{\rho}(Z_{1i}, Z_{2j}|Z_3)$ be the ij-element of the estimated partial correlation matrix of (Z_1, Z_2) given Z_3. This depends on scalar coordinates (z_1, z_2) and a vector coordinate z_3 and assume that this is a continuous function of a vector of local correlations according to one of the alternatives (i) - (v). Then $\alpha^*_{n,b}(i,j|Z_3) - \alpha_{n,b}(i,j|Z_3) \xrightarrow{\mathbb{P}^*_\omega, \mathbb{P}} 0$.

Proof: Denote by $\rho = \rho(Z)$ the vector of local correlations according to one of the alternatives (i) - (v). This vector has finitely many components, and equation (x) 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.

Note: The continuity condition in the assumptions seems not necessary as this follows from the fact that the partial correlation is just an algebraic or rational function of the components of the correlation vector.

The next results gives the asymptotic normality of the bootstrapped local partial correlation. We let $\alpha(i,j)|3) = g_{ij}(\rho(Z)))$ 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)$'$-(v). Using the notation of pages 4 and 5 of this note the correlation vector $\rho(Z)$ will in general have elements consisting of $\rho^{1,2}_{ij}, i = 1, \ldots, d_1, j = 1, \ldots, d_2$; $\rho^{1,3}_{ij}, i = 1, \ldots, d_1, j = 1, \ldots, d_3$; $\rho^{2,3}_{ij}, i = 1, \ldots, d_2, j = 1, \ldots, d_3$, where d_1, d_2, d_3 are chosen according to one of the alternatives (i)$'$ - (v).

Corollary B: Let the assumptions of Theorem B hold and let C be the covariance matrix of the $\hat{\rho}$ vector. Then

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

Proof: As stated above, a proof of Theorem B is given in the proof of Theorem D2 in the supplement of *LT19*. This proof can be adapted to prove Corollary B. 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 $\rho^{1,2}_{ij}$, $\rho^{1,3}_{ij}$, $\rho^{2,3}_{ij}$ that are included in the ρ-vector. The derivations in the proof of Theorem D2 in *LT19*, deal with the more difficult 5-parameter case. In the present case we deal with a sequence of one parameter ρ_{ij}-cases, where cf. *OT17*, 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 *LT19* can be carried through for L^* and L. The only difference is that the quantities $J_{n,b}$ and$M_{n,b}$ now has to be replaced by $J_{n.b}(a) = \mathbb{E}(-\sum_{i,j} a_{ij}\nabla^2 L_{ij,n,b}(\theta_{0,b}))$ and $\mathrm{var}(\sqrt{nb_1b_2}\sum_{i,j} a_{ij}\nabla L_{ij,n,b})$. This results in joint asymptotic bootstrap normality using the Cramer-Wold device. Applying the delta rule now proves the Corollary.

Let M_{ij} and J_{ij} be the M-matrix and J-matrix appearing in the proof of Corollary B for a fixed ρ_{ij}. Note that in the one-parameter case of $\theta_{ij} = \rho_{ij}$ in Corollary B, to first order the covariance matrix C 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 3.1 and Theorem 3.2 of *OT18*. 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 *OT19*.

Asymptotics of testing

General case: $Z = [Z_1, Z_2, Z_3]$ where Z_i has dimension d_i, $i = 1, 2, 3$, and Z has dimension $d_1+d_2+d_3$. We would like to look at the partial correlation matrix of $Z_1, Z_2|Z_3$. In alternative (iii) on page 4 this matrix is given by equation (****) 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 Z_3 by testing whether $\alpha_{i,j} = 0$, or one can test for conditional independence between Z_1 and Z_2 given Z_3 by testing whether $\boldsymbol{\alpha} = \{\alpha_{ij}\} = \mathbf{0}$. The latter case is 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 *LT18* we write, letting $z = [z_1, z_2, z_3]$,

$$T = \int_S h(\rho(z))dF(z)$$

where S is an a area of integration, F is the cumulative distribution function, and $\rho = \rho(z)$, as before, is a vector consisting of all possible pairs that can be formed by picking pairwise components from Z. The number of components of this vector depends on which of the alternatives models (i) - (v) on page 5-6 one considers. It is reasonable to express T as a function of ρ since all partial correlations can be expressed as a function of this vector (see e.g. again equation (****) on page 4.). One possible choice of h is $h(\rho(z)) = \sum_{i,j}\alpha_{i,j}^2(z_{1,i}, z_{2,j}|z_3)$ which is a function of ρ since the α-s are functions of ρ. The corresponding function in *LT18* is $h(\theta(x)$, where $\theta(x)$ is at most 5-dimensional and x is bivariate. As in that paper we proceed in two stages. First we let the bandwidth vector b be fixed and let $n \to \infty$, and then let $b \to 0$. Accordingly we define

$$T_b = \int_S h(\rho_b(z))dF(z)$$

where ρ_b is the population quantity for a fixed b. And then $T_{n,b}$ by

$$T_{n,b} = \int_S h(\rho_{b,n}(z))dF(z)$$

where $\rho_{b,n}(z)$ is the estimates of $\rho(z)$ obtained by pairwise likelihood for each component of $\rho(z)$.

Consistency is easy to prove:

Theorem (Consistency of test). Let the conditions of Theorem(Consistency) on page 9 be fulfilled. Then $T_{n,b} \xrightarrow{\mathbb{P}} T_b$ as $n \to \infty$ and $T_{n,b} \xrightarrow{\mathbb{P}} T$ as $n \to \infty$ and $b \to 0$.

Proof: By the continuous mapping theorem.

To obtain the asymptotic distribution is harder. We will follow the technique of proof in the proof of Theorem 3.1 in *LT18*. See also Joe (1989). The details are a bit more involved because of the larger dimension and because one needs 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 one needs to look at the estimates $\rho_{ij,n,b}(z_i, z_j)$ for all possible pairs (i, j) of the ρ vector, and b) the case where $d_1 = d_2 = d_3 = 1$, but where $\rho_{ij} = \rho_{ij}(z)$ is allowed to depend on all three coordinates $z = (z_1, z_2, z_3)$. We will put the emphasis on a) and only comment briefly on b).

Theorem C: Let the assumptions of Theorem 3 of *OT18* be fulfilled. Then as $n \to \infty$

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

Here

$$A_b(z) = \int_S \Big[\sum_{i,j} \{\frac{\partial}{\partial \rho_{ij}} h(\rho_b, (v)\} J_{ij}^{-1}(\rho_{ij,b}(v_{ij}))K_b(z_{ij} - v_{ij})u(z_{ij}, \rho_b(v_{ij})\Big]dF(v)$$

$$+h(\rho_b(z))1_S(z)$$

and

$$C_n(A_b) = \int A_b^2(z)dF(z) + \frac{1}{n}\sum_{k \neq l} \int A_b(z)A_b(u)dF^{k,l}(z, u) - n\int A_b(z)A_b(u)dF(z)dF(u) \quad (t0)$$

with $F^{k,l}(z,u)$ being the joint cumulative distribution function of (Z^k, Z^l), and where $\{Z^i, i = 1, \ldots, n\}$ are the n observations of Z. Moreover, to ease comparison with *LT18* we write $z_{ij} \doteq (z_i, z_j)$ and $v_{ij} \doteq (v_i, v_j)$. 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. Further if $n \to \infty$ and $b_n \to 0$, the relationship between them being as in Assumption 3 of Theorem 3.1 in *OT19*, then T_b can be replaced by T,

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

Proof: From the proof of Theorem 3.1 in *LT18*, one has that asymptotically,

$$T_{n,b} - T_b \sim \frac{1}{\sqrt{n}}\Big\{\int_S h(\rho_b(z))dG_n(z) + \int_S \nabla_\rho h(\rho_b(z)^T V_n(z) dF(z)\Big\}. \qquad (t1)$$

Here

$$G_n(z) = \sqrt{n}(F_n(z) - F(z)) \quad \text{and} \quad V_n(z) = \sqrt{n}((\rho_{n,b}(z) - \rho_b(z))$$

with F_n being the empirical cumulative function, and where again ρ 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 Z^k, $k = 1, \ldots, n$ so that a mixing CLT can be used. We proceed to the pairwise analysis.

Let (i,j) be a fixed pair and $\rho_{ij,n,b}(z_i, z_j)$ be the corresponding estimated local correlation for the pair (i,j) chosen from the vector $\rho_{n,b}$. Here the pair of coordinates $z_{ij} \doteq (z_i, z_j)$ corresponds to the bivariate x in *LT18*, proof of Theorem 3.1. 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

$$L_{ij,n} = \frac{1}{n}\sum_{k=1}^{n} K_b(Z_{ij}^k - z_{ij}) \log \psi(Z_{ij}^k, \rho_{b,ij}(z_{ij})) - \int K_b(v_{ij} - z_{ij}) \psi_{ij}(v_{ij}, \rho_{b,ij}(z_{ij})) dv_{ij}$$

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 $Z_{ij}^k = (Z_i^k, Z_j^k)$ is the k-th pair of observations, and $v_{ij} = (v_i, v_j)$. Estimates $\rho_{ij,b,n}(z_{ij})$ are obtained by differentiation of the log likelihood and by introducing the score function U_{ij} and information function I_{ij} as follows

$$U_{ij}(z_{ij}, \rho_{ij}) = \frac{\partial}{\partial \rho_{ij}} L_{ij,n}(z_{ij}, \rho_{ij})$$

$$= \frac{1}{n}\sum_{k=1}^{n} K_b(Z_{ij}^k - z_{ij}) u(Z_{ij}^k, \rho_{ij}) - \int K_b(v_{ij} - z_{ij}) u(v_{ij}, \rho_{ij}) \psi(v_{ij}, \rho_{ij}) dv_{ij}$$

where $u(z_{ij}, \rho_{ij}) = \frac{\partial}{\partial \rho_{ij}} \log \psi(z_{ij}, \rho_{ij})$. Similarly

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

where *LT18* may be consulted for more details in the derivation and a more explicit expression for $I_{ij}(z_{ij}, \rho_{ij})$ for the case considered there. In contradistinction to *LT18*, $J_{b,ij}$ is not singular as $b \to 0$. In fact, it is shown in *OT18* that as $b \to 0$, $J_{b,ij} \to u(z_{ij}, \rho_{ij})\psi(z_{ij}, \rho(z_{ij}))$ The vector $V_n(z) = \sqrt{n}[\rho_{n,b}(z) - \rho_b(z)]$ has the dimension of $\rho(z)$ and consists of components $\{\rho_{ij,n,b}(z_{ij}) - \rho_{ij,b}(z_{ij})\}$ where i and j in the most general case run over all possible pairs formed by the components of $\rho(z)$. To be able to use the mixing CLT we follow *LT18* and express this more explicitly in terms of the variables Z^k, $k = 1, \ldots, n$. By Taylor expansion as in *LT18*,

$$0 = U(Z_{ij}, \rho_{n,b}(z_{ij})) \sim U(Z_{ij}, \rho_b(z_{ij})) + I(Z_{ij}, \rho_b(z_{ij}))(\rho_{n,b}(z_{ij}) - \rho_b(z_{ij}))$$
$$\sim U(Z_{ij}, \rho_b(z_{ij})) - J_{ij,b}(z_{ij})(\rho_{n,b}(z_{ij}) - \rho_b(z_{ij}))$$

which implies

$$\rho_{n,b}(z_{ij}) - \rho_b(z_{ij}) \sim J_{ij,b}^{-1}(z_{ij})U(Z_{ij}, \rho_b(z_{ij}))$$

$$= J_{ij,b}^{-1}(z_{ij})[\frac{1}{n}\sum_{k=1}^{n} K_b(Z_{ij}^k - z_{ij})u(Z_{ij}^k, \rho_b(z_{ij})) - \int K_b(v_{ij} - z_{ij})u(v_{ij}, \rho_b(z_{ij}))\psi(v_{ij}, \rho_b(z_{ij}))dv_{ij}] \tag{t2}$$

The second term in (t1) can then be expressed as

$$\int_S \sum_{i,j} \frac{\partial}{\partial \rho_{ij}} h(\rho_b(z)) \times$$

$$J_{ij,b}^{-1}(z_{ij})[\frac{1}{n}\sum_{k=1}^{n} K_b(Z_{ij}^k - z_{ij})u(Z_{ij}^k, \rho_b(z_{ij})) - \int K_b(v_{ij} - z_{ij})u(v_{ij}, \rho_b(z_{ij}))\psi(v_{ij}, \rho_b(z_{ij}))dv_{ij}]dF(z) \tag{t2}$$

where the sum over (i, j) extends over all possible ρ_{ij} contained in the correlation vector ρ. This should be put together with the first term of (t1)

$$\frac{1}{\sqrt{n}} \int_S h(\rho_b(z))dG_n(z) = \frac{1}{n}\sum_{k=1}^{n} h(\rho_b(Z^k)) - \int h(\rho_b(z)dF(z) \tag{t3}$$

We have now reached an expression where one can easily apply the mixing CLT, and an application of that theorem to the sum in (t2) and (t3) now yields asymptotic normality of $\sqrt{n}(T_{n,b} - T_b)$.

Using the same argument as in *LT18* separately on the two main terms yields zero means. Moreover, using the definitions of G_n and $A_b(z)$ it is seen from (t2) and (t3) that

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

By analogy with the proof of Theorem 3.1 in *LT18* it is seen that

$$\mathrm{var}(\int A_b(z)dG_n(z)) \sim \int A_b^2(z)dF(z) + \frac{1}{n}\sum_{k \neq l} \int A_b(z)A_b(u)dF^{k,l}(z, u)$$

$$-n\int A_b(z)A_b(u)dF(z)dF(u) = C_n(A_b)$$

which is the expression (t0) 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 *LT18*.

As in *LT19*, proof of Theorem 3.1, the last part follows straightforwardly from Prop. 6.3.9 of Brockwell and Davis. CHECK!

It should be carefully noted that it is not entirely obvious what the order of the term $C_n(A_b)$ is as $b \to 0$ and $n \to \infty$. Looking at the expressions (t2) and (t3), it is the term (t2) that causes potential trouble due to the presence of b_1^{-1} and b_2^{-1} in the two kernel function terms in the product kernel. However these are canceled out by the integration over (z_1, z_2) in (t2). 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 $\rho_{n,b}(z) - \rho(z)$ is $(nb_1b_2)^{-1/2}$, but this increase in convergence rate is a well-known phenomenon for functionals of type $T_{n,d}$. In *LT19* the convergence rate is increased 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).

CHECK! This variance can be evaluated in more detail as has been done in *LT18*, in particular using the mixing inequality and Hall and Heyde (1980, Corollary A2). But as shown in Teräsvirta et al. (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 in the beginning of this section is the alternative with $d_1 = d_2 = d_3 = 1$ but with correlations depending on three coordinates, so that $z = (z_1, z_2, z_3)$. This is easier in the sense that there are only three pair of correlations, $\rho_{12}, \rho_{13}, \rho_{23}$ and in the above notation $\rho(z) = [\rho_{12}(z), \rho_{13}(z), \rho_{23}(z)]$. The partial correlation is

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

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

$$T_{n,b} = \int_S h(\rho_{b,n}(z))dF(z), \quad T_b = \int_S h(\rho_b(z))dF(z) \;\; T = \int_S h(\rho(z))dF(z)$$

Theorem D: Let the assumptions of Theorem 3 of *OT18* be fulfilled. Then as $n \to \infty$

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

Here

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

and

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

with $F^{k,l}(z,u)$ being the joint cumulative distribution function of (Z^k, Z^l). Further if $n \to \infty$ and $b_n \to 0$, the relationship between them being as in Assumption 3 of Theorem 3.1 in *OT19*, 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 C. Corresponding to (t1)

$$T(\rho_{b,n}) - T(\rho_b) \sim \frac{1}{\sqrt{n}}\Big\{\int_S h(\rho_b(z))dG_n(z) + \int_S \nabla_\rho h(\rho_b(z)))^T V_n(z)\Big\}dF(z)$$

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

$$V_n(z) = \sqrt{n}(\rho_{n,b}(z) - \rho_b(z)) \sim$$

$$\mathbf{J}_b^{-1}(z)\Big[\frac{1}{n}\sum_{k=1}^{n} K_b(Z^k - z)u(Z^k, \rho_b(z)) - \int K_b(v-z)u(v,\rho_b(z))\psi(v,\rho_b(z))dv\Big]$$

Here K_b consists of a triple product of kernels $\prod_{i=1}^3 \frac{1}{b_i}K(\cdot/b_i)$, $u(z,\rho_b(z)) = \nabla_\rho \log\psi(z,\rho_b(z))$ 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 $\rho_b(z)$. This is further developed as in the proof of Theorem C, and the asymptotic normality of $T_{n,b} - T_b$ follows as before from the mixing assumptions. The asymptotic normality of $T_{n,b} - T$ again follows by applying proposition 6.3.9 of Brockwell and Davis.

The order of $C_n(A_b)$ is of interest as $b \to 0$ and $n \to \infty$ since it determines the convergence rate of $T_{n,b} - T$. From Theorem 3.1 of *OT19* it follows that $\rho_{n,b}(z)$ converges at a rate of $(nb^5)^{-1/2}$ to $\rho(z)$ as $n \to \infty$ and $b_1 = b_2 = b_3 = b \to 0$. In the proof of Theorem D 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 C, but this is not unexpected since the faster rate in Theorem C comes at the price of a paiwise simplification in the model structure.

Asymptotic power and Pitman efficiency

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

We will follow the approach used in Wang and Hong (2017, Theorem 2) to derive the results on local asymptotic efficiency. We will explicitly use the results in Theorem C and D with $h(\boldsymbol{\alpha}_{n,b}(z) = ||\boldsymbol{\alpha}_{n,b}(z)||^2$, so that $T_{n,b} = \int_S ||\boldsymbol{\alpha}_{n,b}(z)||^2 dF_n(z)$. The null hypothesis of conditional independence can then be phrased as

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

against the Pitman alternative

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

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 $\beta(z) = \beta(\rho(z))$ is a positive function of z such that $\int_S \beta(z) dF(Z) \doteq \gamma < \infty$. Here $\beta(z)$ is the result of taking an arbitrary correlation vector $\rho(z) = \rho_H(z)$ under the alternative hypothesis such that $\boldsymbol{\alpha}(\rho_H(z)) \neq \mathbf{0}$ and $\beta(z)$ is the Froebenius norm squared of this $\boldsymbol{\alpha}$. Further, $C_{H,n,b}(A_b)$ is the $C_{n,b}$-function formed as in Theorem C from $\rho_H(z)$. It converges towards $C_H(A_b)$ as depicted in the mixing expression of the proof of Theorem C. (uniform convergence needed here?) .

Theorem C$'$: Let the conditions be as in Theorem C and let $H[a_n]$ be as above with $a_n = n^{-1/2}$. Let $C(A)$ be the limit of $C_n(A_b)$ as $n \to \infty$ and $b \to 0$ at the rates specified above. (This limit exists using the mixing expressions of the proof of Theorem C). 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 $\mathbb{P}[T_{n,b} \geq z_\varepsilon | H[a_n]) \to 1 - \Phi(z_\varepsilon - \gamma/[C_H(A))]^{1/2})$ as $n \to \infty, b \to 0$, 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:

$$T_{n,b} = \int_S ||\boldsymbol{\alpha}(z)_{n,b}||^2 dF_n(z)$$

$$= \int_S ||\boldsymbol{\alpha}(z)_{n,b}||^2 - a_n \int_S \beta(z) dF(z) + a_n \int \beta(z) dF(z)$$

Here, using Theorem C, under $H[a_n]$,

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

and

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

and this completes the proof. CHECK!

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

Theorem D$'$: Let the conditions be as in Theorem D 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 $n \to \infty$ and $b \to 0$ at the rates specified above. (This limit exists using the mixing expressions of the proof of Theorem C). 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 $\mathbb{P}[T_{n,b} \geq z_\varepsilon | H[a_n]) \to 1 - \Phi(z_\varepsilon - \gamma/[C_H(A)')]^{1/2}$ as $n \to \infty, b \to 0$, 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

CHECK 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) on page 5).

*** Can now follow up with a comparison and discussion of other results in the literature***.

The bootstrap of the testing statistic $T_{n,b}$.

PRINCIPLE: Replace F_n-quantities with F_n^*-quantities and F-quantities with F_n-quantities, the latter being known under the bootstrap operation. We treat the pairwise model case as in (iii) on page 5 first.

Theorem C^*: Let the assumptions of Theorem C be fulfilled. Then as $n \to \infty$

$$\sqrt{n}[C_n(A_b)]^{-\frac{1}{2}}(T^*_{n,b} - T_{n,b}) \overset{d_{\mathbb{P}^*_\varphi}}{\to} \mathcal{N}(0,1).$$

Here

$$A_b(z) = \int_S \Big[\sum_{i,j} \{ \frac{\partial}{\partial \rho_{ij,n,b}} h(\rho_{n,b}(v) J_{ij}^{-1}(\rho_{ij,n,b}(v_{ij})^{-1} K_b(z_{ij} - v_{ij}) u(z_{ij}, \rho_{ij,n,b}(v_{ij}) \Big] dF(v)$$

$$+ h(\rho_{ij,n,b}(z) 1_S(z)$$

and

$$C_n(A_b) = \int A_b^2(z) dF_n(z) + \frac{1}{n} \sum_{k \neq l} \int A_b(z) A_b(u) dF_n^{k,l}(z,u) - n \int A_b(z) A_b(u) dF_n(z) dF_n(u) \tag{t00}$$

with $F_n^{k,l}(z,u)$ being the empirical joint cumulative distribution function of (Z^k, Z^l). 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. CHECK carefully if the empirical distribution functions can be replaced by theoretical cumulative distribution functions as in *LT*. Can second part of theorem C be used here?

Proof: From proof of Theorem 3.1 in *LT18*, one has that asymptotically,

$$T^*_{n,b} - T_{n,b} \sim \frac{1}{\sqrt{n}}\Big\{\int_S h(\rho_b(z))dG^*_n(z) + \int_S \nabla_\rho h(\rho_b(z)^T \tilde{V}_n(z)dF_n(z)\Big\}. \quad (t11)$$

Here

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

with F^*_n being the empirical cumulative function for the bootstrapped time series, and where again ρ 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 $Z^{k,*}$, $k = 1, \ldots, n$ 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 $\rho^*_{n,b}$. Here the pair of coordinates (z_i, z_j) corresponds to the bivariate x in *LT18*, proof of Theorem 3.1. To ease notation and the comparison with *LT18* we write $z_{ij} \doteq (z_i, z_j)$. 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

$$L^*_{ij,n} = \frac{1}{n}\sum_{k=1}^{n} K_b(Z^{k,*}_{ij} - z_{ij}) \log\psi(Z^{k,*}_{ij}, \rho_{ij,n,b}(z_{ij})) - \int K_b(v_{ij} - z_{ij})\psi_{ij}(v_{ij}, \rho_{ij,n,b}(z_{ij}))dv_{ij}$$

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 $Z^{k,*}_{ij} = (Z^{k,*}_i, Z^{k,*}_j)$ is the k-th pair of bootstrap observations. Estimates $\rho^*_{ij,n,b}(z_{ij})$ are obtained by differentiation of the log likelihood and by introducing the score function U^*_{ij} and information function I^*_{ij} as follows

$$U^*_{ij}(z_{ij}, \rho_{ij,n}) = \frac{\partial}{\partial \rho_{ij,n}} L^*_{ij,n}(z_{ij}, \rho_{ij,n})$$

$$= \frac{1}{n}\sum_{k=1}^{n} K_b(Z^{k,*}_{ij} - z_{ij})u(Z^{k,*}_{ij}, \rho_{ij,n}) - \int K_b(v_{ij} - z_{ij})u(v_{ij}, \rho_{ij,n})\psi(v_{ij}, \rho_{ij,n})dv_{ij}$$

where $u(z_{ij}, \rho_{ij,n}) = \frac{\partial}{\partial \rho_{ij,n}} \log\psi(z_{ij}, \rho_{ij,n})$. Similarly,

$$I^*_{ij}(z_{ij}, \rho_{ij,n}) = \frac{\partial^2}{\partial \rho^2_{ij,n}} L^*_{ij,n}(z_{ij}, \rho_{ij,n}) \quad \text{and} \quad J^*_{ij}(z_{ij}) = \mathbb{E}^*(-I^*_{ij}(z_{ij}, \rho_{ij,n})),$$

where *LT18* may be consulted for more details in the derivation and a more explicit analog expression for $I^*_{ij}(z_{ij}, \rho_{ij,n}$ for the case considered there. In contradistinction to *LT18*, $J_{ij,n,b}$ is not singular as $b \to 0$. In fact, it can be shown that, cf. *OT17*, as $b \to 0$, $J_{ij,n,b} \to u(z_{ij}, \rho_{ij,n})\psi(z_{ij}, \rho(z_{ij,n}))$

CHECK! The vector $\tilde{V}_n(z) = \sqrt{n}[\rho^*_{n,b}(z) - \rho_{n,b}(z)]$ has the dimension of $\rho(z)$ and consists of components $\{\rho^*_{ij,n,b}(z_{ij} - \rho_{ij,n,b}(z_{ij})\}$ where i and j in the most general case run over all possible pairs formed by the components of $\rho(z)$. To be able to use the mixing CLT analogous to Gonçalves and White (2002) we follow *LT18* and express this more explicitly in terms of the variables $Z^{k,*}$, $k = 1, \ldots, n$. By Taylor expansion as in *LT18*,

$$0 = U^*(z_{ij}, \rho^*_{n,b}(z_{ij})) \sim U^*(z_{ij}, \rho_{b,n}(z_{ij})) + I^*(z_{ij}, \rho_{b,n}(z_{ij}))(\rho^*_{n,b}(z_{ij}) - \rho_{b,n}(z_{ij}))$$

$$\sim U^*(z_{ij}, \rho_{b,n}(z_{ij})) - J_{ij,n,b}(z_{ij})(\rho^*_{n,b}(z_{ij}) - \rho_{n,b}(z_{ij}))$$

where we have used that $\mathbb{E}^*$ is the mean value under the measure $\mathbb{P}^*_\omega$ and reasoning analogous to Theorem 3.1 in *LT18*. This implies

$$\rho^*_{n,b}(z_{ij}) - \rho_{b,n,ij}(z_{ij}) \sim J^{-1}_{ij,n,b}(z_{ij})u(Z^*_{ij}, \rho_{n,b}(z_{ij}))$$

$$= J^{-1}_{ij,n,b}(z_{ij}[\frac{1}{n}\sum_{k=1}^{n} K_b(Z^{k,*}_{ij} - z_{ij})u(Z^{k,*}_{ij}, \rho_{n,b}(z_{ij}))$$

$$- \int K_b(v_{ij} - z_{ij})u(v_{ij}, \rho_{n,b}(z_{ij}))\psi(v_{ij}, \rho_{n,b}(z_{ij})dv_{ij}] \qquad (t22)$$

The second term in (t11) can then be expressed as

$$\int_S \sum_{i,j} \frac{\partial}{\partial \rho_{ij,n,b}} h(\rho_{n,b}(z)) \times J^{-1}_{ij,n,b}(z_{ij})[\frac{1}{n}\sum_{k=1}^{n} K_b(Z^{k,*}_{ij} - z_{ij})u(Z^{k,*}_{ij}, \rho_{n,b}(z_{ij})$$

$$- \int K_b(v_{ij} - z_{ij})u(v_{ij}, \rho_{n,b}(z_{ij}))\psi(v_{ij}, \rho_{n,b}(z_{ij})dv_{ij}]dF_n(z) \qquad (t22)$$

where the sum over (i, j) extends over all ρ_{ij} contained in ρ-vector. This should be put together with the first term of (t11),

$$\frac{1}{\sqrt{n}}\int_S h(\rho_{n,b}(z))dG^*_n(z) = \frac{1}{n}\sum_{k=1}^{n} h(\rho_{n,b}(Z^{k,*}) - \int h(\rho_{n,b}(z)dF_n(z) \qquad (t33)$$

and we have reached an expression where one can apply the mixing CLT as in Gonçalves and White and an application of their result, quoted as Theorems C.1 and E.1 in the supplementary materials of *LT18* and *LT19*, to the expressions in (t22) and (t33).This yields asymptotic normality of $\sqrt{n}(T^*_{n,b} - T_{b,n})$.

To make more clear what are the variables $\{Y_t\}$ and $\{Y^*_t\}$ in Theorem C.1 of *LT18* (and $\{Z^{(b)}_t\}$, $\{(Z^{(b)}_t)^*\}$ of Theorem E.1 of *LT19* , let $\{Y_t\} = A_b(Z_t)\}$ and $\{Y^*_t = A_b(Z^*_t)\}$. Using the same argument as in *LT18* and *LT19* and using the definitions of G^*_n and $A_b(z)$ it is seen from (t22) and (t33) that

$$\sqrt{n}(T_{n,b} - T_b) \sim \int_S A_b(z)dG^*_n(z)$$

By analogy with the proof of Theorem 3.1 in LT18 it is seen that $\mathbb{E}^*(\int A_b(z)dG_n^*(z)) = 0$, and

$$\mathrm{var}^*(\int A_b(z)dG_n^*(z)) \sim \int A_b^2(z)dF_n(z) + \frac{1}{n}\sum_{i\neq k}\int A_b(z)A_b(u)dF_n^{i,k}(z,u)$$

$$-n\int A_b(z)A_b(u)dF_n(z)dF_n(u) = C_n(A_b)$$

which proves the theorem. The sum in the second term of $C_n(A_b)$ in (t00) taken together with the last term converges due to the mixing condition and the mixing inequality of Hall and Heyde (1980, Corollary A2). Details of this argument are identical to those of the last part of the proof of Theorem 3.1 in *LT18*.

As in *LT18*, proof of Theorem 3.1, the last part follows straightforwardly from Prop. 6.3.9 of Brockwell and Davis. CHECK!

A Theroem D^* can now be formulated analogously.

Theorem D^* Let the assumptions of Theorem D be fulfilled. Then as $n \to \infty$ and $b \to 0$

$$\sqrt{n}[C_n(A_b)]^{-\frac{1}{2}}(T^*_{n,b} - T_{n,b}) \stackrel{d_{\mathbb{P}^*}}{\to} \mathcal{N}(0, I).$$

Here

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

and

$$C_n(A_b) = \int A_b^2(z)dF_n(z)+\frac{1}{n}\sum_{k\neq l}\int A_b(z)A_b(u)dF_n^{k,l}(z,u)-n\int A_b(z)A_b(u)dF_n(z)dF_n(u)$$

with $F_n^{k,l}(z,u)$ being the empirical joint cumulative distribution function of (Z^i, Z^k).

Testing and the local bootstrap

The limiting results for $\alpha_n(z)$ (or $\boldsymbol{\alpha}(z)$ in the matrix case) and $T_{b,n}$, and the corresponding bootstrap results can be used to construct asymptotic and bootstrap confidence intervals. In that sense they give an indication of whether $\alpha(z) = 0$ locally or $\alpha(z) \equiv 0$ globally. But such confidence intervals cannot in general be used to test $\alpha(z) \equiv 0$, that is conditional independence, with a given level of the test. To be able to do that one must device 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 for at least two reasons: 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 the probability 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 in Paparoditis and Politis (2000) and much more extensively by Su and White (2008). It has also been used (without proof) in Wang and Hong (2017) and in 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 nevertheless in this supplementary material we will 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 (2017).

In our notation Z_{3t} is the basic process corresponding to X_t in Su and White (2008). Under H_0 we assume that Z_{1t} and Z_{2t} are conditionally independent given Z_{3t}, where Z_{1t} and Z_{2t} correspond to Y_t and Z_t in Wang and Hong (2008).

The first step consists in constructing a bootstrap realization $\{Z_{3t}^*\}$ of $\{Z_{3t}\}$ by sampling from the kernel density estimate $\tilde{f}(z_3) = \frac{1}{n}\sum_{t=1}^{n} L_b(z_3 - Z_{3t})$, where K_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 $\{Z_{1t}^*\}$ and $\{Z_{2t}^*\}$ are then created independently of each other by sampling from the conditional kernel estimators $\tilde{f}(z_1|Z_{3t}^*) = \sum_s L_b(Z_{1s} - z_1)L_b(Z_{3s} - Z_{3t}^*)/\sum_s L_b(Z_{3s} - Z_{3t}^*)$ and $\tilde{f}(z_2|Z_{3t}^* = \sum_s L_b(Z_{2s} - z_2)L_b(Z_{3s} - Z_{3t}^*)/\sum_s L_b(Z_{3s} - Z_{3t}^*)$, 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 $\{Z_{1t}^*, Z_{2t}^*, Z_{3t}^*\}$ 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 $\{Z_t\}$, then the level-ε critical values $\tilde{c}_\varepsilon$ are computed as approximate solutions of $\mathbb{P}^*(T_n^* > \tilde{c}_\varepsilon) = \varepsilon$ and the bootstrap p-value is given by $p^* = B^{-1}\sum_{j=1}^{B} 1(T_{nj}^* > T_n)$, 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 $\{Z_{1t}^*, Z_{2t}^*, Z_{3t}^*\}$. The bootstrap realization $\{Z_{3t}^*\}$ can be created using the ordinary bootstrap on $\{Z_{3t}\}$, cf. Paparoditis and Politis (2000), or the kernel smoothed bootstrap as in Su and White

(2008). If the dimension of Z_{31t} is large a smoothed bootstrap realization can be created by sampling from the local Gaussian density estimate of *OT17* using say acceptance reject sampling. In the next step in creating the conditional bootstrap realizations $\{Z_{2t}^*\}$ and $\{Z_{3t}^*\}$ 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 *OT18*. Again it should be noted that the avoidance of the kernel estimate comes at the cost of more restrictive model. We compared the two types of estimates in *OT18*.

Creating bootstrap realizations by the above scheme is quite time consuming since for each Z_{3t}^*, drawings of Z_{1t}^* and Z_{2t}^* 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 by....... Håkon please fill in!

Once a bootstrap realization $\{Z_{1t}^*, Z_{2t}^*, Z_{3t}^*\}$ is created, following Su and White (2018), these realizations are now taken as the original observations in the local likelihood estimation procedure. For a fixed bandwidth h 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. *OT18*, *OT17*. Note that in this paper we have largely operated with two alternatives, the general vector case with inference via the pairwise local likelihood, alternative (iii) on page 5, or the the purely scalar case with three coordinates for the local Gaussian correlation, alternative (i) on page 5. In both cases we have normalized the processes so that they have standard normal marginals. Finally, note that it is recommended; Su and White (2008), that the bandwidth b in creating the bootstrap realizations should be larger than the bandwidth h. In the Monte Carlo experiments of Su and White (2008), it seems that $b = n^{-1/5}$ and $h = n^{-1/3}$.

In the $\{Z_{1t}^*, Z_{2t}^*, Z_{3t}^*\}$-framework we obtain, following the notation in Su and White (2018), an estimate $\hat{\rho}^*$ 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 *OT17*, and subsequently an asymptotic theory for the test statistic T_n^* using the method of proof in Theorem 3.1 in LT18 and Theorem 3.2 in LT19. See also the formulation of those proofs in the present paper, theorems C* and D*, where alternative (iii) and (i) above are accommodated. The scaling factor $C_n(a_b)$ is the same as in the formulation of those results with obvious notational changes. Moreover under the present null hypotheses $T_h = 0$, leading to the asymptotic results that the scaled T_n^*, $T_n^* \xrightarrow{d} \mathcal{N}(0, 1)$.

Note that an alternative proof may possibly be constructed by using the general theory for the smoothed bootstrap as applied to statistical functionals. See Wang, S. (1995). "Optimizing the smooth bootstrap". Annals of the Institute of Statistical Mathematics 47, 65-80. Quite a lot was written about the properties of the smoothed bootstrap in the 80s and 90s following an early paper by Efron. Two of these papers are Silverman and Young (1987). "The bootstrap: to smooth or not to smooth?", Biometrika, 469-479. de Angelis and Young (1992). "Smoothing the bootstrap". International Statistical Institute.

*** A GENERAL POINT: The primary purpose of the paper is to introduce a local Gaussian version of the partial correlation function. This reduces to the ordinary partial correaltion in the Gaussian case. It comes with a sign giving a direction to conditional dependence. The primary purpose is not to obtain a new test of conditional dependence although this is an obvious consequence of the construction of the local partial correlation. It or other tests can be used to test for conditional independence. If the null hypothesis of conditional dependence is rejected, the local Gaussian partial correlation can be used to quantify the strenth and direction of conditional dependence. Our test of conditional independence starts from a perhaps stricter model than other test. On the other hand it is more robust towards the curse of dimensionality. Despite its somewhat restrictive model structure it seems to perform on par with the best existing tests. In addition it has potential for improvements, for example by allowing a 5-parameter pairwise structure model instead of a one-parameter one for the Z-variables. A 5-parameter Z-structure was used in JT19 in their study of nonlinear spectral anaysis. The 5-parameter approach gave substantial improvements over the one-parameter one. ***