# Nonlinear spectral analysis: A local Gaussian approach

Lars Arne Jordanger* Dag Tjøstheim†

This is an Accepted Manuscript of an article published by Taylor & Francis in Journal of the American Statistical Association. (Accepted author version posted online: 26 Oct 2020, Published online: 04 Jan 2021.) Available online: https://www.tandfonline.com/doi/10.1080/01621459.2020.1840991.

**Abstract**

The spectral distribution $f(\omega)$ of a stationary time series $\{Y_t\}_{t\in\mathbb{Z}}$ can be used to investigate whether or not periodic structures are present in $\{Y_t\}_{t\in\mathbb{Z}}$, but $f(\omega)$ has some limitations due to its dependence on the autocovariances $\gamma(h)$. For example, $f(\omega)$ can not distinguish white i.i.d. noise from GARCH-type models (whose terms are dependent, but uncorrelated), which implies that $f(\omega)$ can be an inadequate tool when $\{Y_t\}_{t\in\mathbb{Z}}$ contains asymmetries and nonlinear dependencies.

Asymmetries between the upper and lower tails of a time series can be investigated by means of the *local Gaussian autocorrelations* introduced in Tjøstheim and Hufthammer [2013], and these *local measures of dependence* can be used to construct the *local Gaussian spectral density* presented in this paper. A key feature of the new local spectral density is that it coincides with $f(\omega)$ for Gaussian time series, which implies that it can be used to detect non-Gaussian traits in the time series under investigation. In particular, if $f(\omega)$ is flat, then peaks and troughs of the new local spectral density can indicate nonlinear traits, which potentially might discover *local periodic phenomena* that remain undetected in an ordinary spectral analysis.

*Keywords:* Local periodocities, GARCH models, graphical tools.

*Western Norway University of Applied Sciences, Faculty of Engineering and Science, P.B 7030, 5020 Bergen, Norway E-mail: Lars.Arne.Jordanger@hvl.no

†University of Bergen, Department of Mathematics, P.B. 7803, 5020 Bergen, Norway

# 1 Introduction

Spectral analysis is an important tool in time series analysis. In its classical form, assuming $\sum |\gamma(h)| < \infty$, the spectral density function of a stationary times series $\{Y_t\}_{t\in\mathbb{Z}}$ is the Fourier transform of the autocovariances $\big\{\gamma(h) = \mathrm{Cov}\big(Y_{t+h}, Y_t\big)\big\}_{h\in\mathbb{Z}}$. Furthermore, since $\gamma(h) = \mathrm{Var}(Y_t) \cdot \rho(h)$, with $\rho(h)$ the autocorrelations, this can be expressed as:

$$f(\omega) := \sum_{h\in\mathbb{Z}} \gamma(h) \cdot e^{-2\pi i\omega h} = \mathrm{Var}(Y_t) \cdot \sum_{h\in\mathbb{Z}} \rho(h) \cdot e^{-2\pi i\omega h}. \tag{1}$$

The connection $\mathrm{Var}(Y_t) = \int_{-1/2}^{1/2} f(\omega)\,\mathrm{d}\omega$ follows from the inverse Fourier transformation, and this reveals how $f(\omega)$ gives a decomposition of the variance over different frequencies. In particular, the spectral density function $f(\omega)$ captures the components of periodic linear structure decomposed over frequency for $\{Y_t\}_{t\in\mathbb{Z}}$, and the peaks and troughs of $f(\omega)$ can thus reveal important features of the time series under investigation.

Nonlinear dependencies between the terms of a time series $\{Y_t\}_{t\in\mathbb{Z}}$ will however not be reflected in the spectral density $f(\omega)$, since it is the linear dependencies that are detected by the autocovariance functions $\gamma(h)$. The most obvious example is the GARCH model from Bollerslev [1986]. The GARCH model is much used in econometrics, and it is well known that this model in general exhibits dependence over many lags (long range dependence). But this dependence is not captured by the autocovariance function, since $\gamma(h)$ is zero for lags $|h| \geq 1$. This again implies that the spectral density is flat for a GARCH model.

An estimate of $f(\omega)$ based on samples from, for example, a GARCH(1,1)-model will then, as seen in the left panel of Figure 1, not reveal any information at all. An investigation based on the method presented in this paper can however detect the nonlinear structure — as seen in the right panel of Figure 1, where a point in the lower tail has been inspected.

One may ask whether there exist classes of processes for which the spectral density gives complete information about the probabilistic dependence structure. The answer is simple:

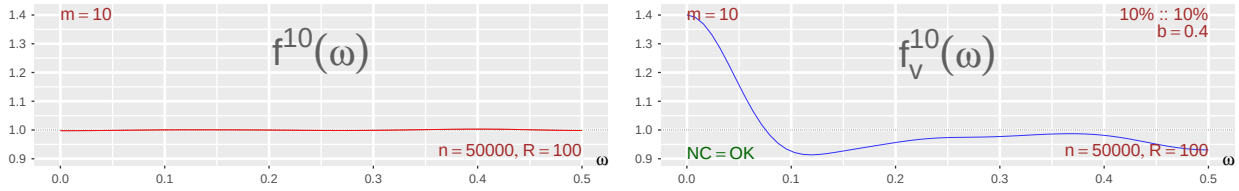


Figure 1: Left: Estimated ordinary (variance-rescaled) spectral density based on a GARCH(1,1)-example. Right: Estimated local Gaussian spectral density at a point in the lower tail. See Appendix G.2 for details regarding the underlying data.

If $\{Y_t\}_{t\in\mathbb{Z}}$ is a stationary Gaussian process, then its complete distributional dependence structure (assuming a zero mean process) can be set up in terms of its spectral density. This is in fact a starting point for the Whittle-type likelihood in time series analysis.

This paper is concerned with finding a generalization of Equation (1) that enables the investigation of nonlinear structures in general non-Gaussian stationary processes. This will be based on a local approach using Gaussian approximations, which ensures the desirable property that the ordinary spectral density is returned for a Gaussian process.

A number of attempts have been made in the literature to extend the standard spectral density $f(\omega)$, and these can roughly be divided into three categories.

Perhaps the best known, and probably the procedure going furthest back in time, is represented by the higher order spectra; see Brillinger [1984, 1991]; Tukey [1959]. The formula for the ordinary spectral density $f(\omega)$ from Equation (1) is then supplemented by considering the Fourier transformations of the higher order moments (or cumulants), such as $\mathrm{E}[Y_rY_sY_t]$ resulting in the bispectrum depending on a double set of frequencies and $\mathrm{E}[Y_rY_sY_tY_u]$ producing the trispectrum dependent on a triple of frequencies. These cumulant-based higher order spectra are identical to zero for Gaussian processes. The multi-

frequency dependence of the bispectrum and trispectrum is not always easy to interpret, and one may also question the existence of higher order moments; in econometrics thick tails often makes this into an issue.

Another approach is to replace $\gamma(h)$ in Equation (1) by another measure of dependence as a function of $h$. Recently there has been much activity in constructing an alternative to Equation (1) by considering covariances of a stationary process obtained by describing quantile crossings, see Hagemann [2011] for a well-written introduction and many references. This is a local spectrum in the sense that it varies with the chosen quantile. It is not always possible to give a local periodic frequency interpretation as in Equation (1), but Li [2012c] emphasized a local sinusoidal construction by analogy with quantile regression models. See also Han et al. [2016]; Li [2008, 2010a,b,c, 2012a,b, 2014]; Linton and Whang [2007]. These approaches does usually not recover the ordinary spectrum for the Gaussian processes. This loss of recovery is also the case if a local spectrum is constructed on the basis of the so-called conditional correlation function (Silvapulle and Granger [2001]). Still another viewpoint would be obtained in a spectral analysis of the distance Brownian covariance function Székely and Rizzo [2009].

A third alternative is constituted by Hong's generalized spectrum, see Hong [1999, 2000], which is obtained by replacing the covariance function $\gamma(h)$ in Equation (1) by the bivariate covariance function $\sigma_h(u,v)$ constructed by taking covariances between the characteristic function expressions $\exp\left(iuY_{t+h}\right)$ and $\exp\left(ivY_h\right)$. Again, this gives a complete distributional characterization of dependence properties, but so far not much attention has been given to concrete data analytic interpretation of this frequency representation. Rather, it has been used to test for independence, conditional independence and predictability Li et al. [2016]; Wang and Hong [2018].

The new approach presented in this paper follows the strategy where the $\gamma(h)$ of Equation (1) is replaced by another dependence measure, that is, the *local Gaussian autocorrelation* introduced in Tjøstheim and Hufthammer [2013], see Lacal and Tjøstheim [2017, 2018] for a number of recent references. The definition of the local Gaussian autocorrelation is given in section 2.1, but the gist of it can be described as follows: The joint distribution of $\left(Y_{t+h}, Y_t\right)$ is approximated locally at a point $\boldsymbol{v}$, say, by a Gaussian bivariate distribution — and the correlation parameter from this approximating Gaussian distribution is then taken as the local Gaussian autocorrelation $\rho_{\boldsymbol{v}}(h)$ at the point $\boldsymbol{v}$. If $\sum |\rho_{\boldsymbol{v}}(h)| < \infty$, the *local Gaussian spectral density* at the point $\boldsymbol{v}$ can be defined in the following manner,

$$f_{\boldsymbol{v}}(\omega) \coloneqq \sum_{h=-\infty}^{\infty} \rho_{\boldsymbol{v}}(h) \cdot e^{-2\pi i \omega h}. \tag{2}$$

This enables a local frequency decomposition with different frequency representations at different points $\boldsymbol{v}$, for example, different oscillatory behaviour at extremes (cf. also the extremogram of Davis and Mikosch [2009]) as compared to oscillatory behaviour in the center of the process. The point $\boldsymbol{v}$ will naturally correspond to a pair of quantiles, but this concept is distinctly different from the quantile spectra referred to above in that it considers a neighbourhood of $\boldsymbol{v}$ and not $\boldsymbol{v}$ as a threshold. Moreover, this approach returns a scaled version of the ordinary spectrum when a Gaussian process is investigated, with equality when $\mathrm{Var}(Y_t) = 1$.

Due to issues related to numerical convergence, the estimates presented in this paper will be based on an initial normalization of $\{Y_t\}_{t\in\mathbb{Z}}$, and for the normalized processes the correlation $\rho(h)$ will always equal the covariance $\gamma(h)$. All references to $f(\omega)$ will henceforth refer to the spectral density of a normalized process, that is, $f(\omega)$ will now refer to the

following rescaled version instead of the one given in Equation (1),

$$f(\omega) := \sum_{h \in \mathbb{Z}} \rho(h) \cdot e^{-2\pi i \omega}. \tag{3}$$

For the normalized processes, $f(\omega)$ and $f_{\boldsymbol{v}}(\omega)$ will by construction be identical for Gaussian time series, and a comparison of the ordinary spectrum $f(\omega)$ and the local Gaussian spectrum $f_{\boldsymbol{v}}(\omega)$ can thus be used to investigate at a local level how a non-Gaussian time series deviates from being Gaussian.

Much more details of this framework is given in section 2. This section also contains the asymptotic theory with detailed proofs in the Supplementary Material. The real and simulated examples of section 3 show that local spectral estimates can detect local periodic phenomena and detect nonlinearities in non-Gaussian white noise. Note that the scripts needed for the reproduction of these examples are contained in the `R`-package `localgaussSpec`,[1] where it in addition is possible to use an interactive tool to see how adjustments of the input parameters (used in the estimation algorithms) influence the estimates of $f_{\boldsymbol{v}}(\omega)$.

The theory developed in this paper can be extended to the multivariate case, see Jordanger and Tjøstheim [2017] .

## 2 Local Gaussian spectral densities

The local Gaussian correlation (LGC) was introduced in Tjøstheim and Hufthammer [2013], with theory that showed how it could be used to estimate the local Gaussian autocorrelations for a time series. It has been further developed in a number of papers, primarily Lacal and Tjøstheim [2017, 2018], but see also Berentsen et al. [2017, 2014a]; Berentsen and Tjøstheim [2014]; Berentsen et al. [2014b]; Otneim and Tjøstheim [2017, 2018]; Støve

[1] Use `remotes::install_github("LAJordanger/localgaussSpec")` to install the package. See the documentation of the function `LG_extract_scripts` for further details. See also Appendix G.

and Tjøstheim [2014]; Støve et al. [2014] for related issues. In Tjøstheim and Hufthammer [2013] the possibility of developing a local Gaussian spectral analysis was briefly mentioned, and this is the topic of the present paper.

This section gives a brief summary of the local Gaussian autocorrelations, and use them to define the local Gaussian spectral density for strictly[2] stationary univariate time series $\{Y_t\}_{t\in\mathbb{Z}}$, and give estimators with a corresponding asymptotic theory.

## 2.1 The local Gaussian correlations

Details related to the estimation regime, and asymptotic properties, can be found in Appendix B.1.2 in the Supplementary Material. Note that other approaches to the concept of local Gaussian correlation also have been investigated, cf. Berentsen et al. [2017] for details.

### 2.1.1 Local Gaussian correlation, general version

Consider a bivariate random variable $\boldsymbol{W} = (W_1, W_2)$ with joint cdf $G(\boldsymbol{w})$ and joint pdf $g(\boldsymbol{w})$. For a specified point $\boldsymbol{v} \coloneqq (v_1, v_2)$, the main idea is to find the bivariate Gaussian distribution whose density function best approximates $g(\boldsymbol{w})$ in a neighbourhood of the point of interest. The LGC will then be defined to be the correlation of this local Gaussian approximation.

For the purpose of this investigation, the vector containing the five local parameters $\mu_1$, $\mu_2$, $\sigma_1$, $\sigma_2$ and $\rho$ will be denoted by $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{v})$,[3] and the approximating bivariate Gaussian density function at the point $\boldsymbol{v}$ will be denoted $\psi(\boldsymbol{w}; \boldsymbol{\theta})$, that is,

$$\psi(\boldsymbol{w}; \boldsymbol{\theta}) \coloneqq \frac{1}{2\pi\cdot\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \exp\left\{-\frac{\sigma_1^2\big(w_1-\mu_1\big)^2 - 2\sigma_1\sigma_2\rho\big(w_1-\mu_1\big)\big(w_2-\mu_2\big) + \sigma_2^2\big(w_2-\mu_2\big)^2}{2\sigma_1^2\sigma_2^2(1-\rho^2)}\right\}. \tag{4}$$

[2] Strict stationarity is necessary in order for the machinery of the local Gaussian approximations to be feasible, since Gaussian pdfs will be used to locally approximate the pdfs corresponding to the bivariate pairs $\big(Y_{t+h}, Y_t\big)$.

[3] The vector $\boldsymbol{\theta}$ is a function of the point $\boldsymbol{v}$, but this will henceforth be suppressed in the notation.

In order for $\psi(\boldsymbol{w};\boldsymbol{\theta})$ to be considered a good approximation of $g(\boldsymbol{w})$ in a neighbourhood of the point $\boldsymbol{v}$, it should at least coincide with $g(\boldsymbol{w})$ at $\boldsymbol{v}$, and it furthermore seems natural to require that the tangent planes should coincide too, that is,

$$g(\boldsymbol{v}) = \psi(\boldsymbol{v};\boldsymbol{\theta}), \tag{5a}$$

$$\frac{\partial}{\partial w_1} g(\boldsymbol{v}) = \frac{\partial}{\partial w_1} \psi(\boldsymbol{v};\boldsymbol{\theta}) \text{ and } \frac{\partial}{\partial w_2} g(\boldsymbol{v}) = \frac{\partial}{\partial w_2} \psi(\boldsymbol{v};\boldsymbol{\theta}). \tag{5b}$$

It is easy to verify analytically that a solution $\boldsymbol{\theta}$ can be found for any point $\boldsymbol{v}$ where $g(\boldsymbol{w})$ is smooth — but these solutions are not unique: $\psi(\boldsymbol{w};\boldsymbol{\theta})$ and $\psi(\boldsymbol{w};\boldsymbol{\theta}')$ can have the same first order linearization around the point $\boldsymbol{v}$, without $\boldsymbol{\theta}$ being identical to $\boldsymbol{\theta}'$. It is tempting to extend Equation (5) to also include similar requirements for the second order partial derivatives, but the system of equations will then in general have no solution.

This shows that it, in order to find the local Gaussian parameters in $\boldsymbol{\theta}$, is insufficient to only consider requirements at $\boldsymbol{v}$, it is necessary to apply an argument that also takes into account a neighbourhood around $\boldsymbol{v}$. Applying the approach used when estimating densities in Hjort and Jones [1996], one can consider a $\boldsymbol{b} \to \boldsymbol{0}^+$ limit of parameters $\boldsymbol{\theta}_{\boldsymbol{b}} = \boldsymbol{\theta}_{\boldsymbol{b}}(\boldsymbol{v})$ that minimize the penalty function

$$q_{\boldsymbol{b}} = \int K_{\boldsymbol{b}}(\boldsymbol{w} - \boldsymbol{v}) \left[\psi(\boldsymbol{w};\boldsymbol{\theta}) - g(\boldsymbol{w}) \log\left(\psi(\boldsymbol{w};\boldsymbol{\theta})\right)\right] \mathrm{d}\boldsymbol{w}, \tag{6}$$

where $K_{\boldsymbol{b}}(\boldsymbol{w} - \boldsymbol{v})$ is a kernel function with bandwidth $\boldsymbol{b}$. As explained in Hjort and Jones [1996, Section 2.1], this can be interpreted as a locally weighted Kullback-Leibler distance between the targeted density $g(\boldsymbol{w})$ and the approximating density $\psi(\boldsymbol{w};\boldsymbol{\theta})$. An optimal parameter configuration $\boldsymbol{\theta}_{\boldsymbol{b}}$ for Equation (6) should solve the vector equation

$$\int K_{\boldsymbol{b}}(\boldsymbol{w} - \boldsymbol{v})\, \boldsymbol{u}(\boldsymbol{w};\boldsymbol{\theta}) \left[\psi(\boldsymbol{w};\boldsymbol{\theta}) - g(\boldsymbol{w})\right] \mathrm{d}\boldsymbol{w} = \boldsymbol{0}, \tag{7}$$

where $\boldsymbol{u}(\boldsymbol{w};\boldsymbol{\theta}) \coloneqq \frac{\partial}{\partial \boldsymbol{\theta}} \log\left(\psi(\boldsymbol{w};\boldsymbol{\theta})\right)$ is the score function of the approximating density $\psi(\boldsymbol{w};\boldsymbol{\theta})$. There will, under suitable assumptions Hjort and Jones [1996]; Tjøstheim and Hufthammer

[2013], be a unique limiting solution of Equation (7), that is,

$$\boldsymbol{\theta}_0 = \boldsymbol{\theta}_0(\boldsymbol{v}) = \lim_{\boldsymbol{b} \to \mathbf{0}^+} \boldsymbol{\theta}_{\boldsymbol{b}}(\boldsymbol{v}) \tag{8}$$

will be well-defined,[4] and the $\rho$-part of the $\boldsymbol{\theta}_0$-vector can be used to define a LGC at the point $\boldsymbol{v}$.

For the special case where $g(\boldsymbol{w})$ is a bivariate normal distribution, that is, when

$$\boldsymbol{W} \sim \mathrm{N}\left( \begin{bmatrix} \mu_1 \\ \mu_1 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{bmatrix} \right), \tag{9}$$

then, for any point $\boldsymbol{v}$ and any bandwidth $\boldsymbol{b}$, the parameters $\boldsymbol{\theta}_{\boldsymbol{b}}$ that gives the optimal solution of Equation (7) will be the parameters given in Equation (9). The limit $\boldsymbol{\theta}_0$ in Equation (8) will thus of course also be these parameters, which implies that the LGC coincides with the global parameter $\rho$ at all points in the Gaussian case. The interested reader should consult Tjøstheim and Hufthammer [2013, p. 33] for further details/remarks that motivates the use of the LGC.

An estimate of the local Gaussian parameters $\boldsymbol{\theta}_0(\boldsymbol{v})$ in Equation (8) can, for a given bivariate sample $\{\boldsymbol{W}_t\}_{t=1}^n$ and some reasonable bandwidth $\boldsymbol{b}$, be found as the parameter-vector $\hat{\boldsymbol{\theta}}_{\boldsymbol{b}}(\boldsymbol{v})$ that maximizes the local log-likelihood[5]

$$L_n(\boldsymbol{\theta}) \coloneqq n^{-1} \sum_{t=1}^{n} K_{\boldsymbol{b}}(\boldsymbol{W}_t - \boldsymbol{v}) \log \psi(\boldsymbol{W}_t; \boldsymbol{\theta}) - \int_{\mathbb{R}^2} K_{\boldsymbol{b}}(\boldsymbol{w} - \boldsymbol{v})\, \psi(\boldsymbol{w}; \boldsymbol{\theta})\, \mathrm{d}\boldsymbol{w}. \tag{10}$$

The asymptotic behaviour of $\hat{\boldsymbol{\theta}}_{\boldsymbol{b}}(\boldsymbol{v})$ (as $n \to \infty$ and $\boldsymbol{b} \to \mathbf{0}^+$) is in Tjøstheim and Hufthammer [2013] investigated by entities derived from a local penalty function $Q_n(\boldsymbol{\theta})$ defined as $-n \cdot L_n(\boldsymbol{\theta})$, that is,

$$Q_n(\boldsymbol{\theta}) = -\sum_{t=1}^{n} K_{\boldsymbol{b}}(\boldsymbol{W}_t - \boldsymbol{v}) \log \psi(\boldsymbol{W}_t; \boldsymbol{\theta}) + n \int_{\mathbb{R}^2} K_{\boldsymbol{b}}(\boldsymbol{w} - \boldsymbol{v})\, \psi(\boldsymbol{w}; \boldsymbol{\theta})\, \mathrm{d}\boldsymbol{w}. \tag{11}$$

[4] The solution $\boldsymbol{\theta}_0$ will always satisfy Equation (5a), but it will in general not satisfy Equation (5b).

[5] Confer Appendix B.1.2 in the supplementary material for a detailed exposition.

The key ingredient in the analysis is the corresponding vector of partial derivatives,

$$\nabla Q_n(\boldsymbol{\theta}) = -\sum_{t=1}^{n}\left[K_{\boldsymbol{b}}(\boldsymbol{W}_t - \boldsymbol{v})\,\boldsymbol{u}(\boldsymbol{W}_t;\boldsymbol{\theta}) - \int_{\mathbb{R}^2} K_{\boldsymbol{b}}(\boldsymbol{w}-\boldsymbol{v})\,\boldsymbol{u}(\boldsymbol{w};\boldsymbol{\theta})\,\psi(\boldsymbol{w};\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{w}\right], \quad (12)$$

and, as will be seen later on, the asymptotic investigation of the local Gaussian spectral density $f_{\boldsymbol{v}}(\omega)$ introduced in this paper does also build on this entity.

Notice that the bias-variance balance of the estimate $\hat{\boldsymbol{\theta}}_{\boldsymbol{b}}(\boldsymbol{v})$ depends on the bandwidth-vector $\boldsymbol{b}$, and an estimate based on a $\boldsymbol{b}$ too close to $\mathbf{0}$ might thus be dubious. However, it can still be of interest (for a given sample) to compare estimates $\hat{\boldsymbol{\theta}}_{\boldsymbol{b}}(\boldsymbol{v})$ for different scales of $\boldsymbol{b}$ in order to see how they behave.

Since the goal is to estimate $\boldsymbol{\theta}_0(\boldsymbol{v})$, it is of course important to find $\hat{\boldsymbol{\theta}}_{\boldsymbol{b}}(\boldsymbol{v})$ for not too large bandwidth-vectors $\boldsymbol{b}$ — but it might still be of interest to point out how Equation (10) behaves in the "global limit $\boldsymbol{b} \to \boldsymbol{\infty} = (\infty, \infty)$". In this case the second term goes to zero, and the parameter-vector $\hat{\boldsymbol{\theta}}_{\boldsymbol{\infty}}(\boldsymbol{v})$ that maximizes the first term becomes the ordinary (global) least squares estimates of a global parameter vector $\boldsymbol{\theta}$ which contains the ordinary means, variances and correlation.

### 2.1.2 Local Gaussian correlation, normalized version

The algorithm that estimates the LGC (see Berentsen and Tjøstheim [2014] for an `R`-implementation) can run into problems if the data under investigation contain outliers — that is, the numerical convergence might not succeed for points $\boldsymbol{v}$ in the periphery of the data. It is possible to counter this problem by removing the most extreme outliers, but an alternative strategy based on normalization will be applied instead.

The key observation is that the numerical estimation problem does not occur when the marginal distributions are standard normal — which motivates an adjusted strategy similar to the copula-concept from Sklar [1959]. Sklar's theorem gives the existence of a copula $C(u_1, u_2)$ such that the joint cdf $G(\boldsymbol{w})$ can be expressed as $C(G_1(w_1), G_2(w_2))$,

with $G_i(w_i)$ the marginal cdf corresponding to $W_i$. This copula $C$ contains all the interdependence information between the two marginal random variables $W_1$ and $W_2$, it will be unique when the two margins are continuous, and it will then be invariant under strictly increasing transformations of the margins.[6] Under this continuity assumption, the random variable $\boldsymbol{W} = (W_1, W_2)$ will have the same copula as the transformed random variable $\boldsymbol{Z} \coloneqq (\Phi^{-1}(G_1(W_1)), \Phi^{-1}(G_2(W_2)))$, where $\Phi$ is the cdf of the standard normal distribution — whose corresponding pdf as usual will be denoted by $\phi$.[7] This transformed version of $\boldsymbol{W}$ has standard normal margins, so the LGC-estimation algorithm will not run into numerical problems — which motivates the following alternative approach to the definition of LGC: Instead of finding a Gaussian approximation of the pdf $g(\boldsymbol{w})$ (of the original random variable $\boldsymbol{W}$) at a point $\boldsymbol{v}$, find a Gaussian approximation of the pdf $g_{\boldsymbol{Z}}(\boldsymbol{z})$ of the transformed random variable $\boldsymbol{Z}$ at a transformed point $\boldsymbol{v}_{\boldsymbol{Z}}$. Expressed relative to the pdf $c$ of the copula $C$, this means that the setup in Equation (13b) below will be used instead of the setup in Equation (13a).

$$g(\boldsymbol{w}) = c(G_1(w_1), G_2(w_2))\, g_1(w_1)\, g_2(w_2) \quad \text{approximate at } \boldsymbol{v} = (v_1, v_2), \tag{13a}$$

$$g_{\boldsymbol{Z}}(\boldsymbol{z}) = c(\Phi(z_1), \Phi(z_2))\, \phi(z_1)\, \phi(z_2) \quad \text{approximate at } \boldsymbol{v}_{\boldsymbol{Z}} \coloneqq \left(\Phi^{-1}(G_1(v_1)), \Phi^{-1}(G_2(v_2))\right). \tag{13b}$$

The normalized version of the LGC will return values that differ from those obtained from the general LGC-version introduced in section 2.1.1, but the two versions coincide when the random variable $\boldsymbol{W}$ is bivariate Gaussian. The transformed random variable $\boldsymbol{Z}$ corresponding to the $\boldsymbol{W}$ from Equation (9) will then be $\boldsymbol{Z} = ((W_1 - \mu_1)/\sigma_1, (W_2 - \mu_2)/\sigma_2)$,

[6] For a proof of this statement, see, for example, Nelsen [2006, Theorem 2.4.3].

[7] See Berentsen et al. [2014b] for an approach where this is used to construct a *canonical local Gaussian correlation* for the copula $C$.

which implies

$$\boldsymbol{Z} \sim \mathrm{N}\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right), \tag{14}$$

so the normalized LGC will thus also coincide with the global parameter $\rho$ at all points.

The convergence rate for the estimates is rather slow for the LGC cases discussed above (it is $\sqrt{n(b_1 b_2)^3}$), and that is due to the kernel function $K_{\boldsymbol{b}}$ in Equation (6). Briefly summarized, the $5 \times 5$ covariance matrix of the estimate $\hat{\boldsymbol{\theta}}_{\boldsymbol{b}}$ will have the form $V_{\boldsymbol{b}}^{-1} W_{\boldsymbol{b}} V_{\boldsymbol{b}}^{-1}$, the presence of the kernel $K_{\boldsymbol{b}}$ means that the matrices $V_{\boldsymbol{b}}$ and $W_{\boldsymbol{b}}$ have rank one in the limit $\boldsymbol{b} \to \mathbf{0}^+$, and this slows down the convergence rate, cf. Tjøstheim and Hufthammer [2013, Theorem 3] for the details.

The property that the limiting matrices have rank one does not pose a problem if only one parameter is estimated,[8] and the convergence rate would then be much faster (that is, $\sqrt{nb_1 b_2}$). Inspired by the fact that the transformed random variable $\boldsymbol{Z}$ have standard normal margins, it has been introduced a simplified normalized version of the LGC where only the $\rho$-parameter should be estimated when using the approximation approach from Equation (13b), that is, the values of $\mu_1$, $\mu_2$ are taken to be 0, whereas $\sigma_1^2$ and $\sigma_2^2$ are taken to be 1. This simplified approach has been applied successfully with regard to density estimation[9] in Otneim and Tjøstheim [2017, 2018], but for the local spectrum analysis considered in this paper it gave inferior results — and this paper will thus not include any plots based on the normalized one-parameter version.[10]

[8] The matrices then becomes $1 \times 1$, so the singularity problems does not occur.

[9] Note that it is *not* the local Gaussian correlation that is the target of interest when this simplified approach is used for density estimation.

[10] The theory for the normalized one-free-parameter version of LGC is avaialbe in the first authors PhD-thesis, `https://bora.uib.no/handle/1956/16950`. This also contains a discussion with regard to why an approach based on the normalized one-free-parameter approach fails to produce decent results.

## 2.2 The local Gaussian spectral densities

An extension of the spectral density $f(\omega)$ from Equation (3) can be based on any of the three LGC-versions mentioned in sections 2.1.1 and 2.1.2. The one presented below is based on the normalized five-parameter local Gaussian autocorrelation, since that ensures that the estimation algorithm avoids the aforementioned numerical convergence problems — but the theory developed in the Supplementary Material does also cover the general situation.

**Definition 2.1.** The local Gaussian spectral density (LGSD), at a point, $\boldsymbol{v} = (v_1, v_2)$, for a strictly stationary univariate time series $\{Y_t\}_{t\in\mathbb{Z}}$ is constructed in the following manner.

1. With $G$ the univariate *marginal* cumulative distribution of $\{Y_t\}_{t\in\mathbb{Z}}$, and $\Phi$ the cumulative distribution of the standard normal distribution, define a normalized version $\{Z_t\}_{t\in\mathbb{Z}}$ of $\{Y_t\}_{t\in\mathbb{Z}}$ by
$$\{Z_t \coloneqq \Phi^{-1}(G(Y_t))\}_{t\in\mathbb{Z}}. \tag{15}$$
2. For a given point $\boldsymbol{v} = (v_1, v_2)$ and for each *bivariate pair* $\boldsymbol{Z}_{h:t} \coloneqq \big(Z_{t+h}, Z_t\big)$, a *local Gaussian autocorrelation* $\rho_{\boldsymbol{v}}(h)$ can be computed. The convention $\rho_{\boldsymbol{v}}(0) \equiv 1$ is used when $h = 0$.
3. When $\sum_{h\in\mathbb{Z}} |\rho_{\boldsymbol{v}}(h)| < \infty$, the *local Gaussian spectral density* at the point $\boldsymbol{v}$ is defined as
$$f_{\boldsymbol{v}}(\omega) \coloneqq \sum_{h=-\infty}^{\infty} \rho_{\boldsymbol{v}}(h) \cdot e^{-2\pi i\omega h}. \tag{16}$$

Notice that the requirement $\sum_{h\in\mathbb{Z}} |\rho_{\boldsymbol{v}}(h)| < \infty$ in Definition 2.1(3) implies that the concept of local Gaussian spectral density in general might not be well defined for all stationary time series $\{Y_t\}_{t\in\mathbb{Z}}$ and all points $\boldsymbol{v} \in \mathbb{R}^2$.

The normalization in Equation (15) preserves the copula-structure of the original time series, but a standard normal marginal will be used instead of its original marginal distribution. This implies that the transformed time series will have all moments, even though that might not be the case for a tick tailed original time series. A local Gaussian investigation of the normalized time series can detect non-Gaussian dependency structures in the original time series, but keep in mind that an investigation of the original marginal might also be of interest in many situations, for example, with regard to discriminant analysis.

Finally, note that the normalization in Equation (15) can be compared to, but is very different from, the normalization in Klüppelberg and Mikosch [1994].

The following definition of time reversible time series, from Tong [1990, def. 4.6], is needed in Lemma 2.1(3).

**Definition 2.2.** A stationary time series $\{Y_t\}_{t\in\mathbb{Z}}$ is time reversible if for every positive integer $n$ and every $t_1, t_2, \ldots, t_n \in \mathbb{Z}$, the vectors $\left(Y_{t_1}, Y_{t_2} \ldots, Y_{t_n}\right)$ and $\left(Y_{-t_1}, Y_{-t_2} \ldots, Y_{-t_n}\right)$ have the same joint distributions.

**Lemma 2.1.** *The following properties holds for $f_{\boldsymbol{v}}(\omega)$.*

1. *$f_{\boldsymbol{v}}(\omega)$ coincides with $f(\omega)$ for all $\boldsymbol{v} \in \mathbb{R}^2$ when $\{Y_t\}_{t\in\mathbb{Z}}$ is a Gaussian time series, or when $\{Y_t\}_{t\in\mathbb{Z}}$ consists of i.i.d. observations.*
2. *The following holds when $\breve{\boldsymbol{v}} := (v_2, v_1)$ is the diagonal reflection of $\boldsymbol{v} = (v_1, v_2)$;*

$$f_{\boldsymbol{v}}(\omega) = 1 + \sum_{h=1}^{\infty} \rho_{\breve{\boldsymbol{v}}}(h) \cdot e^{+2\pi i\omega h} + \sum_{h=1}^{\infty} \rho_{\boldsymbol{v}}(h) \cdot e^{-2\pi i\omega h}, \tag{17a}$$

$$f_{\boldsymbol{v}}(\omega) = \overline{f_{\breve{\boldsymbol{v}}}(\omega)}. \tag{17b}$$

3. *When $\{Y_t\}_{t\in\mathbb{Z}}$ is time reversible, then $f_{\boldsymbol{v}}(\omega)$ is real valued for all $\boldsymbol{v} \in \mathbb{R}^2$, that is,*

$$f_{\boldsymbol{v}}(\omega) = 1 + 2 \cdot \sum_{h=1}^{\infty} \rho_{\boldsymbol{v}}(h) \cdot \cos(2\pi\omega h). \tag{18}$$

4. *$f_{\boldsymbol{v}}(\omega)$ will in general be complex-valued, but it will always be real valued when the point $\boldsymbol{v}$ lies on the diagonal, that is, when $v_1 = v_2$. Equation (18) will hold in this diagonal case too.*

*Proof.* Item 1 follows for the Gaussian case since the local Gaussian autocorrelations $\rho_{\boldsymbol{v}}(h)$ by construction coincides with the ordinary (global) autocorrelations $\rho(h)$ in the Gaussian case. Similarly, when $\{Y_t\}_{t\in\mathbb{Z}}$ consists of i.i.d. observations, then both local and global autocorrelations will be 0 when $h \neq 0$, and the local and global spectra both become the constant function 1. Items 2 to 4 are trivial consequences of the diagonal folding property from Lemma C.1, that is, $\rho_{\boldsymbol{v}}(-h) = \rho_{\breve{\boldsymbol{v}}}(h)$, and the definition of time reversibility, see Appendices C.1 and C.2 for details. □

For general points $\boldsymbol{v} = (v_1, v_2)$, the complex valued result of $f_{\boldsymbol{v}}(\omega)$ might be hard to investigate and interpret — but, due to Lemma 2.1(4), the investigation becomes simpler for points on the diagonal. This might also be the situation of most practical interest, since it corresponds to estimating the local spectrum at (or around) a given value of $\{Y_t\}_{t\in\mathbb{Z}}$ — such as a certain quantile for the distribution of $Y_t$. The real valued results $f_{\boldsymbol{v}}(\omega)$ for $\boldsymbol{v}$ along the diagonal can be compared with the result of the ordinary (global) spectral density $f(\omega)$, as given in Equation (3), and this might detect cases where the times series $\{Y_t\}_{t\in\mathbb{Z}}$ deviates from *being Gaussian*. Furthermore, if the global spectrum $f(\omega)$ is flat, then any peaks and troughs of $f_{\boldsymbol{v}}(\omega)$ might be interpreted as indicators of for example, *periodicities at a local level*. This implies that estimates of $f_{\boldsymbol{v}}(\omega)$ might be useful as an exploratory tool, an idea that will be pursued in section 3.

Note that the collection of local Gaussian autocorrelations $\{\rho_{\boldsymbol{v}}(h)\}_{h\in\mathbb{Z}}$ might not be non-negative definite, which implies that both the theoretical and estimated local Gaussian spectral densities might therefore become negative. However, as the artificial process

investigated in Figure 7 shows, the peaks of $f_{\boldsymbol{v}}(\omega)$ still occur at the expected frequencies for the investigated points — which implies that the lack of non-negativity does not prevent this tool from detecting nonlinear structures in non-Gaussian white noise.

The following definition is needed when the discussion later on refers to $m$-truncated versions of the different spectra.

**Definition 2.3.** The $m$-truncated versions $f_{\boldsymbol{v}}^m(\omega)$ and $f^m(\omega)$ of $f_{\boldsymbol{v}}(\omega)$ and $f(\omega)$, for some lag-window function $\lambda_m(h)$, is defined by means of

$$f_{\boldsymbol{v}}^m(\omega) := 1 + \sum_{h=1}^{m} \lambda_m(h) \cdot \rho_{\check{\boldsymbol{v}}}(h) \cdot e^{+2\pi i\omega h} + \sum_{h=1}^{m} \lambda_m(h) \cdot \rho_{\boldsymbol{v}}(h) \cdot e^{-2\pi i\omega h}, \tag{19a}$$

$$f^m(\omega) := \sum_{h=-m}^{m} \lambda_m(h) \cdot \rho(h) \cdot e^{-2\pi i\omega h}. \tag{19b}$$

## 2.3 Estimation

Theoretical and numerical estimates of the ordinary spectral density $f(\omega)$ is typically investigated by means of the fast Fourier transform (FFT) and techniques related to the periodogram. The FFT-approach can not be used in the local case since there is no natural factorization of terms making up a local estimated covariance, but there does exist a pre-FFT approach for the estimation of $f(\omega)$, where a Fourier transform is taken of the estimated autocorrelations after they have been smoothed and truncated by means of some lag-window function — and the pre-FFT approach can be adapted to deal with the estimates of the local Gaussian spectral densities.

**Algorithm 2.1.** *For a sample $\{y_t\}_{t=1}^n$ of size $n$, an $m$-truncated estimate $\hat{f}_{\boldsymbol{v}}^m(\omega)$ of $f_{\boldsymbol{v}}(\omega)$ is constructed by means of the following procedure.*

1. *Find an estimate $\hat{G}_n$ of the marginal cumulative distribution function, and compute the pseudo-normalized observations $\left\{\hat{z}_t := \Phi^{-1}\left(\hat{G}_n(y_t)\right)\right\}_{t=1}^n$ that corresponds to $\{y_t\}_{t=1}^n$.*

2. *Create the lag $h$ pseudo-normalized pairs $\left\{\left(\hat{z}_{t+h}, \hat{z}_t\right)\right\}_{t=1}^{n-h}$ for $h = 1, \ldots, m$, and estimate, both for the point $\boldsymbol{v} = (v_1, v_2)$ and its diagonal reflection $\breve{\boldsymbol{v}} = (v_2, v_1)$, the local Gaussian autocorrelations $\{\hat{\rho}_{\boldsymbol{v}}(h|\boldsymbol{b}_h)\}_{h=1}^m$ and $\{\hat{\rho}_{\breve{\boldsymbol{v}}}(h|\boldsymbol{b}_h)\}_{h=1}^m$, where the $\{\boldsymbol{b}_h\}_{h=1}^m$ is the bandwidths used during the estimation of the local Gaussian autocorrelation for the different lags.*
3. *Adjust Equation (17a) from Lemma 2.1(2) with some lag-window function $\lambda_m(h)$ to get the estimate*

$$\hat{f}_{\boldsymbol{v}}^m(\omega) \coloneqq 1 + \sum_{h=1}^{m} \lambda_m(h) \cdot \hat{\rho}_{\breve{\boldsymbol{v}}}(h|\boldsymbol{b}_h) \cdot e^{+2\pi i\omega h} + \sum_{h=1}^{m} \lambda_m(h) \cdot \hat{\rho}_{\boldsymbol{v}}(h|\boldsymbol{b}_h) \cdot e^{-2\pi i\omega h}. \tag{20}$$

The presence of the kernel $K_{\boldsymbol{b}}(\boldsymbol{w} - \boldsymbol{v})$ in Equation (6) implies that small sample effects can occur when the local Gaussian spectrum $f_{\boldsymbol{v}}(\omega)$ is estimated for some combinations of points $\boldsymbol{v}$ and bandwidths $\boldsymbol{b}$ — and this can in particular be an issue if the points lie in the low density regions corresponding to the tails of our distribution. Roughly speaking: When the bandwidth $\boldsymbol{b}$ becomes "too small", then the estimated local Gaussian autocorrelations will have a tendency to approach either "$-1$" or "$+1$", cf. Appendix D.3 — and these estimates will then in general only reflect the random configuration of those lag-$h$ pairs that happened to lie closest to the point $\boldsymbol{v}$. Section 3.1 presents strategies that can be used in order to detect/avoid this issue, and additional details are presented in the Supplementary Material.

The following result is an analogue to Equation (18) of Lemma 2.1(3)

**Lemma 2.2.** *When it is assumed that the sample $\{y_t\}_{t=1}^n$ comes from a time reversible stochastic process $\{Y_t\}_{t\in\mathbb{Z}}$, the $m$-truncated estimate $\hat{f}_{\boldsymbol{v}}^m(\omega)$ can for all points $\boldsymbol{v} \in \mathbb{R}^2$ be written as*

$$\hat{f}_{\boldsymbol{v}}^m(\omega) = 1 + 2 \cdot \sum_{h=1}^{m} \lambda_m(h) \cdot \hat{\rho}_{\boldsymbol{v}}(h|\boldsymbol{b}_h) \cdot \cos(2\pi\omega h). \tag{21}$$

*Moreover, Equation* (21) *will always hold when the point* $\boldsymbol{v}$ *lies on the diagonal, that is,* $v_1 = v_2$.

*Proof.* This follows from Items 3 and 4 of Lemma 2.1. □

The estimated $\hat{G}_n$ in Algorithm 2.1(2) can for example be the rescaled empirical cumulative distribution function created from the sample $\{y_t\}_{t=1}^n$ (which transforms original data into ranks divided by $n+1$), or it could be based on some logspline technique like the one implemented in Otneim and Tjøstheim [2017].

The bandwidths $\boldsymbol{b}_h = (b_{h1}, b_{h2})$ in Algorithm 2.1(2) does not need to be equal for all the lags $h$ when an estimate $\hat{f}_{\boldsymbol{v}}^m(\omega)$ is computed. For the asymptotic investigation it is sufficient to require that $b_{h1}$ and $b_{h2}$ approach zero at the same rate, that is, that there exists $\boldsymbol{b} = (b_1, b_2)$ such that $b_{hi} \asymp b_i$ for $i = 1, 2$ and for all $h$ (that is to say, $\lim b_{hi}/b_i = 1$).

The asymptotic theory for $\hat{\rho}_{\boldsymbol{v}}(h|\boldsymbol{b}_h)$, given that the required regularity conditions are satisfied, follows when the original argument from Tjøstheim and Hufthammer [2013] is combined with the argument in Otneim and Tjøstheim [2017]. The analysis in Tjøstheim and Hufthammer [2013] considered the general case where the original observations $\{y_t\}_{t=1}^n$ were used instead of the normalized observations $\{z_t \coloneqq \Phi^{-1}(G(y_t))\}_{t=1}^n$. Since the cumulative density function $G$ in general will be unknown, the present asymptotic analysis must work with the pseudo-normalized observations $\{\hat{z}_t\}_{t=1}^n$, which makes it necessary to take into account the difference between the true normalized values $z_t$ and the estimated pseudo-normalized values $\hat{z}_t$. The analysis in Otneim and Tjøstheim [2017] implies that $\hat{G}_n(y_t)$ approaches $G(y_t)$ at a faster rate than the rate of convergence for the estimated local Gaussian correlation, so (under some regularity conditions) the convergence rate of $\hat{\rho}_{\boldsymbol{v}}(h|\boldsymbol{b}_h)$ will thus not be affected by the distinction between $z_t$ and $\hat{z}_t$. The present analysis will

not duplicate the arguments related to this distinction, and the interested reader should consult Otneim and Tjøstheim [2017, sec. 3] for the details.

The bias-variance balance for the estimates $\hat{f}_{\boldsymbol{v}}^{m}(\omega)$ must consider the size of $m$ relative to both $n$ and the bandwidths $\{\boldsymbol{b}_h\}_{h=1}^{m}$, that is, the kernel function reduces the number of observations that effectively contributes to the computations of the estimates — and that number of effective contributors can also depend on the location of the point $\boldsymbol{v}$, that is, whether the point $\boldsymbol{v}$ lies at the center or in the periphery of the pseudo-normalized observations $\left\{\left(\hat{z}_{t+h}, \hat{z}_t\right)\right\}_{t=1}^{n-h}$. Confer section 3.2 for further details.

Figure 2 shows the effect of the pseudo-normalization on the `dmbp` example[11] that will be discussed in section 3.4. The uppermost part shows the original `dmbp`-series (of length 1974) whereas the lowermost part shows the pseudo-normalized transformation of it, and it is clear that the shape of the pseudo-normalized version resembles the shape of the original version.

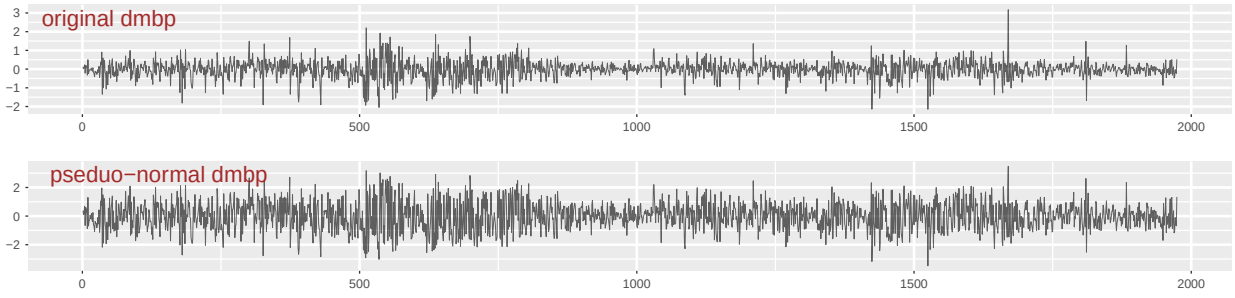


Figure 2: `dmbp`, original version and pseudo-normalized version.

[11] This is the Deutschemark/British pound Exchange Rate (`dmbp`) data from Bollerslev and Ghysels [1996], which is a common benchmark data set for GARCH-type models, and as such models are among the motivating factors for the study of the local Gaussian spectral density, it seems natural to test the method on `dmbp`. The data plotted here was found in the `R`-package `rugarch`, see Ghalanos [2020], where the following description was given: "The daily percentage nominal returns computed as $100\left[\ln\left(P_t\right) - \ln\left(P_t - 1\right)\right]$, where $P_t$ is the bilateral Deutschemark/British pound rate constructed from the corresponding U.S. dollar rates."

## 2.4 Asymptotic theory for $\hat{f}_{\boldsymbol{v}}^{m}(\omega)$

This section presents asymptotic results for the cases where $\hat{f}_{\boldsymbol{v}}^{m}(\omega)$ are real-valued functions. Note that both assumptions and results are stated relative to the original observations instead of the pseudo-normalized observations. This simplification does not affect the final convergence rates (see earlier remarks for details) and it makes the analysis easier. The requirement that the LGSD should be defined relative to the normalized observations is due to computational issues, and the theoretical investigation shows that it could just as well have been phrased in terms of the original observations.

### 2.4.1 A definition and an assumption for $Y_t$

The assumption to be imposed on the univariate time series $\{Y_t\}_{t\in\mathbb{Z}}$ is given in terms of components related to the bivariate lag-$h$-pairs that can be constructed from it. The theoretical analysis of $\hat{f}_{\boldsymbol{v}}^{m}(\omega)$ also requires that $(m+1)$-variate pairs are considered. Note that Item 3 of Algorithm 2.1 implies that it is sufficient to only consider positive values for $h$.

**Definition 2.4.** For a strictly stationary univariate time series $\{Y_t\}_{t\in\mathbb{Z}}$, with $h \geq 1$ and $m \geq 2$, define bivariate and $(m+1)$-variate time series as follows,

$$\boldsymbol{Y}_{h:t} \coloneqq \big[Y_{t+h}, Y_t\big]', \qquad \boldsymbol{Y}_{\overline{m}:t} \coloneqq \big[Y_{t+m}, \ldots, Y_t\big]', \tag{22}$$

and let $g_h(\boldsymbol{y}_h)$ and $g_{\overline{m}}(\boldsymbol{y}_{\overline{m}})$ denote the respective probability density functions.

The bivariate densities $g_h$ can all be obtained from the $(m+1)$ variate density $g_{\overline{m}}$ by integrating out the $m-1$ redundant marginals, which in particular implies that if an $(m+1)$-variate function $\tilde{\eta}_h(\boldsymbol{y}_{\overline{m}}) : \mathbb{R}^{m+1} \to \mathbb{R}^1$ is the *obvious extension*[12] of a bivariate

[12] Consider the function to be a constant with respect to all the new variables that are introduced.

function $\eta_h(\boldsymbol{y}_h) : \mathbb{R}^2 \to \mathbb{R}^1$, then

$$\mathrm{E}[\eta_h(\boldsymbol{Y}_{h:t})] = \mathrm{E}[\tilde{\eta}_h(\boldsymbol{Y}_{\bar{m}:t})]\,, \qquad \text{for } h \in \{1, \ldots, m\}\,. \tag{23}$$

With the notation from Definition 2.4 the following Assumption 2.1 can now be imposed on $Y_t$. Note that Items 5 to 7 of Assumption 2.1 contain references to definitions that first are given explicitly in Appendix B in the Supplementary Material; these definitions are related to an $(m+1)$-variate penalty function for the time series $\boldsymbol{Y}_{\bar{m}:t}$ — and they are quite technical so it would impede the flow of the paper to include all the details here. For the present section, it is sufficient to know that the new $(m+1)$-variate function can be expressed as a sum of $m$ bivariate penalty-functions of the form given in Equation (11).

The key idea is that $\boldsymbol{W}_t$ and $g(\boldsymbol{w})$ in Equations (6) to (12) are replaced with $\boldsymbol{Y}_{h:t}$ and $g_h(\boldsymbol{y}_h)$, which implies that an additional index $h$ must be added in order to keep track of the bookkeeping. In particular, an inspection of Equation (12) motivates the introduction of a random variable vector $\boldsymbol{X}_{h:t} = K_{\boldsymbol{b}}(\boldsymbol{Y}_{h:t} - \boldsymbol{v})\,\boldsymbol{u}(\boldsymbol{Y}_{h:t}; \boldsymbol{\theta})$, and the random variables $X^n_{hq:i}$ that occur in Assumption 2.1(7) are the components of $\sqrt{b_1 b_2}\boldsymbol{X}_{h:t}$. Furthermore, notice that different combinations of the indices $h, i, j$ and $k$ in the product $X^n_{hq:i} \cdot X^n_{jr:k}$ implies that it can contain from two to four different terms of the time series $\{Y_t\}_{t\in\mathbb{Z}}$, so the corresponding density function can thus either be bi-, tri- or tetravariate. The indices $q, r = 1, \ldots, 5$ keep track of the appropriate derivatives of the 5-dimensional parameter vector $\boldsymbol{\theta}$. See Definitions B.7 and B.11 for details.

**Assumption 2.1.** *The univariate process $\{Y_t\}_{t\in\mathbb{Z}}$ will be assumed to satisfy the following properties, with $\boldsymbol{v} = (v_1, v_2)$ in Item 4 the point at which the estimate $\hat{f}^m_{\boldsymbol{v}}(\omega)$ of $f_{\boldsymbol{v}}(\omega)$ is to be computed.*

1. *$\{Y_t\}_{t\in\mathbb{Z}}$ is strictly stationary.*

2. $\{Y_t\}_{t\in\mathbb{Z}}$ *is strongly mixing, with mixing coefficient* $\alpha(j)$ *satisfying*

$$\sum_{j=1}^{\infty} j^a \left[\alpha(j)\right]^{1-2/\nu} < \infty \qquad \text{for some } \nu > 2 \text{ and } a > 1 - 2/\nu. \tag{24}$$

3. $\mathrm{Var}(Y_t) < \infty$.

*The bivariate density functions* $g_h(\boldsymbol{y}_h)$ *of the lag* $h$ *pairs* $\boldsymbol{Y}_{h:t}$ *of the univariate time series* $\{Y_t\}_{t\in\mathbb{Z}}$*, must satisfy the following requirements for a given point* $\boldsymbol{v} = (v_1, v_2)$.

4. $g_h(\boldsymbol{y}_h)$ *is differentiable at* $\boldsymbol{v}$*, such that Taylor's theorem can be used to write* $g_h(\boldsymbol{y}_h)$ *as*

$$g_h(\boldsymbol{y}_h) = g_h(\boldsymbol{v}) + \boldsymbol{\mathfrak{g}}_h(\boldsymbol{v})' \left[\boldsymbol{y}_h - \boldsymbol{v}\right] + \boldsymbol{\mathfrak{R}}_h(\boldsymbol{y}_h)' \left[\boldsymbol{y}_h - \boldsymbol{v}\right], \tag{25}$$

$$\text{where } \boldsymbol{\mathfrak{g}}_h(\boldsymbol{v}) = \left[ \frac{\partial}{\partial y_h} g_h(\boldsymbol{y}_h)\Big|_{\boldsymbol{y}_h=\boldsymbol{v}}, \frac{\partial}{\partial y_0} g_h(\boldsymbol{y}_h)\Big|_{\boldsymbol{y}_h=\boldsymbol{v}} \right]' \text{ and } \lim_{\boldsymbol{y}_h \longrightarrow \boldsymbol{v}} \boldsymbol{\mathfrak{R}}_h(\boldsymbol{y}_h) = \boldsymbol{0},$$

*and the same requirement must also hold for the diagonally reflected point* $\breve{\boldsymbol{v}} = (v_2, v_1)$.

5. *There exists a bandwidth* $\boldsymbol{b}_{h0}$ *such that there for every* $\boldsymbol{0} < \boldsymbol{b} < \boldsymbol{b}_{h0}$ *is a unique minimizer* $\boldsymbol{\theta}_{h:\boldsymbol{b}}$ *of the penalty function* $q_{h:\boldsymbol{b}}$ *defined in Equation* (B.4)*, which is obtained from Equation* (6) *by putting* $\boldsymbol{w} = \boldsymbol{y}_h$.
6. *The collection of bandwidths* $\{\boldsymbol{b}_{h0}\}_{h\in\mathbb{Z}}$ *has a positive infimum, that is, there exists a* $\boldsymbol{b}_0$ *such that* $\boldsymbol{0} < \boldsymbol{b}_0 \coloneqq \inf_{h\in\mathbb{Z}} \boldsymbol{b}_{h0}$*, which implies that this* $\boldsymbol{b}_0$ *can be used simultaneously for all the lags.*
7. *For* $X^n_{hq:i}$ *from Definition* B.11*, the bivariate, trivariate and tetravariate density functions must be such that the expectations* $\mathrm{E}\left[X^n_{hq:i}\right]$, $\mathrm{E}\left[\left|X^n_{hq:i}\right|^{\nu}\right]$ *and* $\mathrm{E}\left[X^n_{hq:i} \cdot X^n_{jr:k}\right]$ *all are finite.*

These assumptions on $Y_t$ are extensions of those used for the LGC-case in Tjøstheim and Hufthammer [2013]. Assumption 2.1(2) is a bit more general than the one used in Tjøstheim and Hufthammer [2013], but that is not a problem since the arguments given there trivially extend to the present case.

The $\alpha$-mixing requirement in Item 2 ensures that $Y_{t+h}$ and $Y_t$ will be asymptotically independent as $h \to \infty$, that is, the bivariate density functions $g_h(\boldsymbol{y}_h)$ will for large lags $h$ approach the product of the marginal densities, and the situation will thus stabilize when $h$ is large enough. This is in particular of importance for Item 6, since it implies that it will be possible to find a nonzero $\boldsymbol{b}_0$ that works for all $h$.

We do not consider the $\alpha$-mixing condition to be very strong. In particular, note that GARCH type models, which are frequently used in econometrics, and also in the present paper, cf. section 3.4.3, are $\beta$-mixing under weak conditions, see, for example, Carrasco and Chen [2002]; and $\beta$-mixing implies $\alpha$-mixing.

The finiteness requirements in Assumption 2.1(7) will be trivially satisfied if the densities are bounded, that is, they will then be consequences of properties of the kernel function $K_{\boldsymbol{b}}$ and the score function of the bivariate Gaussian distribution, see Lemma C.6 for details.

### 2.4.2 An assumption for $Y_t$ and the score function $\boldsymbol{u}(\boldsymbol{w};\boldsymbol{\theta})$ of $\psi(\boldsymbol{w};\boldsymbol{\theta})$

The score function in Equation (7), that is, $\boldsymbol{u}(\boldsymbol{w};\boldsymbol{\theta}) \coloneqq \frac{\partial}{\partial \boldsymbol{\theta}} \log\left(\psi(\boldsymbol{w};\boldsymbol{\theta})\right)$, plays a central role in the local density-estimation approach of Hjort and Jones [1996], and it also plays a pivotal role in the local Gaussian correlation theory developed in Tjøstheim and Hufthammer [2013].

In particular, the convergence rate that in Tjøstheim and Hufthammer [2013] is given for $\hat{\boldsymbol{\theta}}_{\boldsymbol{v}} - \boldsymbol{\theta}_{\boldsymbol{v}}$ does implicitly require that $\boldsymbol{u}(\boldsymbol{v};\boldsymbol{\theta}_{\boldsymbol{v}}) \neq \boldsymbol{0}$ in order for the corresponding asymptotic covariance matrix to be well defined. The investigation of $\left(\hat{f}^m_{\boldsymbol{v}}(\omega) - f_{\boldsymbol{v}}(\omega)\right)$ in this paper builds on the asymptotic results from Tjøstheim and Hufthammer [2013], and the following assumption must be satisfied in order for the given convergence rates and asymptotic variances to be valid.

**Assumption 2.2.** *The collection of local Gaussian parameters* $\{\boldsymbol{\theta}_{\boldsymbol{v}}(h)\}$ *at the point* $\boldsymbol{v}$ *for the bivariate probability density functions* $g_h(\boldsymbol{y}_h)$*, must all be such that*

1. $\boldsymbol{u}(\boldsymbol{v}; \boldsymbol{\theta}_{\boldsymbol{v}}(h)) \neq \mathbf{0}$ *for all finite* $h$*.*
2. $\lim \boldsymbol{u}(\boldsymbol{v}; \boldsymbol{\theta}_{\boldsymbol{v}}(h)) \neq \mathbf{0}$*.*

It is, for a given time series $Y_t$ and a given point $\boldsymbol{v}$, possible to inspect the 5 equations in $\boldsymbol{u}(\boldsymbol{w}; \boldsymbol{\theta}) = \mathbf{0}$ in order to see when Items 1 and 2 of Assumption 2.2 might fail. For the case of the asymptotic requirement in Item 2, the key observation is that the strong mixing requirement from Assumption 2.1(2) implies that $Y_{t+h}$ and $Y_t$ will become independent when $h \to \infty$. Together with the assumption of normalized marginals, this implies that the limit of $\boldsymbol{\theta}_{\boldsymbol{v}}(h)$ always becomes $[\mu_1, \mu_2, \sigma_1, \sigma_2, \rho]' = [0, 0, 1, 1, 0]'$, which means that Assumption 2.2(2) will fail for any point $\boldsymbol{v}$ that solves $\boldsymbol{u}(\boldsymbol{v}; [0, 0, 1, 1, 0]') = 0$.

### 2.4.3 Assumptions for $n$, $m$ and $\boldsymbol{b}$

For simplicity, the present analysis will use the $\boldsymbol{b} = (b_1, b_2)$ introduced in the second paragraph after Lemma 2.2, that is, it will be assumed that the individual bandwidths $\boldsymbol{b}_h$ for the different lags $h$ approach zero at the same rate — and that it for the asymptotic investigation thus can be assumed that the same bandwidth is used for all the lags.

**Assumption 2.3.** *Let* $m \coloneqq m_n \to \infty$ *be a sequence of integers denoting the number of lags to include, and let* $\boldsymbol{b} \coloneqq \boldsymbol{b}_n \to \mathbf{0}^+$ *be the bandwidths used when estimating the local Gaussian correlations for the lags* $h = 1, \ldots, m$ *(based on* $n$ *observations). Let* $b_1$ *and* $b_2$ *refer to the two components of* $\boldsymbol{b}$*, and let* $\alpha$*,* $\nu$ *and* $a$ *be as introduced in Assumption 2.1(2). Let* $s \coloneqq s_n \to \infty$ *be a sequence of integers such that* $s = o\Big(\sqrt{nb_1b_2/m}\Big)$*, and let* $\tau$ *be a positive constant. The following requirements must be satisfied for these entities.*[13]

1. $\log n/n(b_1b_2)^5 \longrightarrow 0$*.*

[13] Notational convention: "$\vee$" denotes the maximum of two numbers, whereas "$\wedge$" denotes the minimum.

2. $nb_1b_2/m \longrightarrow \infty$.
3. $m^\delta(b_1 \vee b_2) \longrightarrow 0$, *where* $\delta = 2 \vee \frac{\nu(a+1)}{\nu(a-1)-2}$.
4. $\sqrt{nm/b_1b_2} \cdot s^\tau \cdot \alpha(s-m+1) \longrightarrow \infty$.
5. $m = o\Big((nb_1b_2)^{\tau/(2+5\tau)-\lambda}\Big)$, *for some* $\lambda \in (0, \tau/(2+5\tau))$.
6. $m = o(s)$.

Assumption 2.3(1) is needed in order for the asymptotic theory from Tjøstheim and Hufthammer [2013] to be valid for the estimates $\hat{\rho}_{\boldsymbol{v}}(h)$. See Lemma C.3 for a verification of the internal consistency of the requirements given in Assumption 2.3. The expected number of observations near $\boldsymbol{v}$ will for large $n$ and small $b_1$ and $b_2$ be of order $nb_1b_2 \cdot g_h(\boldsymbol{v})$ — and this will, when $g_h(\boldsymbol{v}) > 0$, go to infinity when $n \to \infty$ and $\boldsymbol{b} \to \mathbf{0}^+$. See the end of Appendix C.3 for further details.

## 2.5 Convergence theorems for $\hat{f}^m_{\boldsymbol{v}}(\omega)$

**Theorem 2.1** ($\boldsymbol{v}$ on diagonal, that is, $v_1 = v_2$)**.** *The local Gaussian spectral density $f_{\boldsymbol{v}}(\omega)$ is a real valued function when the point $\boldsymbol{v}$ lies on the diagonal. Furthermore; when the univariate time series $Y_t$ satisfies Assumptions 2.1 and 2.2, and $n$, $m$ and $\boldsymbol{b} = (b_1, b_2)$ are as given in Assumption 2.3, then the following asymptotic results holds for the $m$-truncated estimate $\hat{f}^m_{\boldsymbol{v}}(\omega)$,*

$$\sqrt{n(b_1b_2)^3/m} \cdot \Big(\hat{f}^m_{\boldsymbol{v}}(\omega) - f_{\boldsymbol{v}}(\omega)\Big) \xrightarrow{d} \mathrm{N}\big(\mathbf{0}, \sigma^2_{\boldsymbol{v}}(\omega)\big)\,, \tag{26}$$

*where the formula*

$$\sigma^2_{\boldsymbol{v}}(\omega) = 4 \lim_{m\to\infty} \frac{1}{m} \sum_{h=1}^{m} \lambda^2_m(h) \cdot \cos^2(2\pi\omega h) \cdot \tilde{\sigma}^2_{\boldsymbol{v}}(h) \tag{27}$$

*relates the variance $\sigma^2_{\boldsymbol{v}}(\omega)$ to the asymptotic variances $\tilde{\sigma}^2_{\boldsymbol{v}}(h)$ of $\sqrt{n(b_1b_2)^3} \cdot (\hat{\rho}_{\boldsymbol{v}}(h|_{\boldsymbol{b}_h}) - \rho_{\boldsymbol{v}}(h))$.*

*Proof.* The proof is given in Appendix A.1. □

The variance $\sigma^2_{\boldsymbol{v}}(\omega)$ depends on all the bivariate density functions through the variances $\tilde{\sigma}^2_{\boldsymbol{v}}(h)$. Moreover, it is clear from Equation (27) that $\sigma^2_{\boldsymbol{v}}(\omega)$ as a function of the frequency $\omega$ is symmetric around $\omega = \frac{1}{4}$, with its highest values when $\omega \in \left\{0, \frac{1}{2}\right\}$. The same symmetry is not present for the variance of the $m$-truncated spectra $\hat{f}^m_{\boldsymbol{v}}(\omega)$, and the variance of $\hat{f}^m_{\boldsymbol{v}}(\omega)$ will have its highest value when $\omega = 0$, cf. Appendix A.3 for details.

A similar result to Theorem 2.1 can be stated for time reversible stochastic processes.

**Theorem 2.2** ($Y_t$ time reversible)**.** *The local Gaussian spectral density $f_{\boldsymbol{v}}(\omega)$ is a real valued function for all points $\boldsymbol{v}$ when $Y_t$ is time reversible (see Definition 2.2). Furthermore under Assumptions 2.1 to 2.3, the same asymptotic results as stated in Theorem 2.1 holds for the $m$-truncated estimate $\hat{f}^m_{\boldsymbol{v}}(\omega)$.*

*Proof.* Lemma 2.1(3) states that $f_{\boldsymbol{v}}(\omega)$ is a real-valued function, and the proof of Theorem 2.1 (see Appendix A.1) can then be repeated without any modifications. □

The asymptotic normality results in Theorems 2.1 and 2.2 do not easily enable a computation of pointwise confidence intervals for the estimated LGSD. Thus, the pointwise confidence intervals later on will either be estimated based on suitable quantiles obtained by repeated sampling from a known distribution, or they will be based on bootstrapping techniques for those cases where real data have been investigated. Confer Teräsvirta et al. [2010, ch. 7.2.5 and 7.2.6] for further details with regard to the need for bootstrapping in such situations. See also Lacal and Tjøstheim [2017, 2018] for analytic results on the bootstrap and block bootstrap in the case of estimation of the local Gaussian auto- and cross-correlation functions.

The asymptotic result for $\hat{f}_{\boldsymbol{v}}^m(\omega)$ complex-valued is given in Appendix A.2, where it can be seen that $\sqrt{n(b_1 b_2)^3/m} \cdot \left(\hat{f}_{\boldsymbol{v}}^m(\omega) - f_{\boldsymbol{v}}(\omega)\right)$ then asymptotically approaches a complex-valued normal distribution.

# 3 Visualizations and interpretations

This section will show how different visualizations of the $m$-truncated estimates $\hat{f}_{\boldsymbol{v}}^m(\omega)$ can be used to detect nonlinear dependency structures in a time series. Similar graphical methods can also be found in Birr et al. [2019]; Li [2019], and the heatmap-plot presented in this section is in particular inspired by the one encountered in Li [2019].

Technical details, and the description of the selected tuning parameters of $\hat{f}_{\boldsymbol{v}}^m(\omega)$, are given in section 3.1. Section 3.2 uses the aforementioned `dmbp`-data to highlight how the different tuning parameters of the estimation algorithm are interconnected.

A sanity test of the implemented estimation algorithm is presented in section 3.3, and it is there seen that $\hat{f}_{\boldsymbol{v}}^m(\omega)$ can detect local periodic structures in an example where a heuristic argument enables the prediction of the anticipated result. Section 3.4 applies the local Gaussian machinery to the `dmbp`-data, and it also contains the results from a GARCH-type model fitted to the `dmbp`-data. A comparison of the results from the original data and the fitted model can reveal to what extent the internal dependency structure of the fitted model actually reflects the dependency structure of the original sample, and this might be of interest with regard to model selection.

A few extreme examples have been included in the Supplementary Material in order to investigate the limitations of this method. Appendix G.4.3 examine the detection of a periodic component located far out in the tail of a large sample, and Appendix G.4.4 consider a situation based on a deterministic function perturbed by very low random fluctuations.

## 3.1 The input parameters and some other technical details

Several tuning parameters must be selected in order to compute the $m$-truncated local Gaussian spectral density estimates $\hat{f}_{\boldsymbol{v}}^m(\omega)$, and the values used for the plots in this section are given below. Note that these parameters have been selected in order to provide a *proof of concept* for the fact that nonlinear dependency structures can be detected by this approach, and the quest for "optimal parameters" is a topic for further work. The interested reader can consult Appendix D in the Supplementary Material for a sensitivity analysis of the different tuning parameters.

**The pseudo-normalization:** The initial step of the computation of $\hat{f}_{\boldsymbol{v}}^m(\omega)$ is to replace the observations $\{y_t\}_{t=1}^n$ with the corresponding pseudo-normalized observations $\{\hat{z}_t\}_{t=1}^n$, cf. Algorithm 2.1, that is, an estimate of the marginal cumulative density function $G$ is needed. The present analysis has used the rescaled empirical cumulative density function $\hat{G}_n$ for this purpose, but the computations could also have been based on a logspline-estimate of $G$. A preliminary test revealed that the two normalization procedures created strikingly similar estimates of $\hat{f}_{\boldsymbol{v}}^m(\omega)$, so the computationally faster approach based on the rescaled empirical cumulative density-function has thus been applied for the present investigation.

**The length $n$ of the samples:** All samples have the same length as the `dmbp`-data, that is, $n = 1974$. The estimation machinery produces similar results for shorter samples, but it is important to keep in mind that too short samples might not reveal the dependency structure of interest — which in particular might be an issue for the tails of the distribution.

**The points $\boldsymbol{v}$ of investigation:** Three diagonal points, with coordinates corresponding to the 10%, 50% and 90% percentiles of the standard normal distribution,[14] will be used in the basic plots in this section. These points will often be referred to as *lower tail*,

[14] The corresponding coordinates are $(-1.28, -1.28)$, $(0, 0)$ and $(1.28, 1.28)$.

*center* and *upper tail* when discussed in the text. Confer Appendix D.3 for further details related to the selection of $\boldsymbol{v}$, and see Figure 8 for a heatmap-based plot.

**The lag-window function $\lambda_m(h)$:** The smoothing of the estimated local Gaussian autocorrelations, cf. Algorithm 2.1(3), was done by the Tukey-Hanning lag-window kernel: $\lambda_m(h) = \frac{1}{2} \cdot \left(1 + \cos\left(\pi \cdot \frac{h}{m}\right)\right)$ for $|h| \leq m$, $\lambda_m(h) = 0$ for $|h| > m$.

**The bandwidth $\boldsymbol{b}$:** The estimation of the local Gaussian autocorrelations requires the selection of a bandwidth-vector $\boldsymbol{b} = (b_1, b_2)$, and the majority of the plots in this section have used $\boldsymbol{b} = (.5, .5)$. Note that it is natural to require $b_1 = b_2$ since both of the components in the lag $h$ pseudo-normalized pairs comes from the same univariate time series. Further discussion of choice of bandwidth is given in Appendix E

**The truncation level $m$:** The value $m = 10$ was used for the truncation level, since it was possible to detect nonlinear dependency structures even for that low truncation level.

**The number of replicates $R$:** The estimated values (means and 90% pointwise confidence intervals) have been based on $R = 100$ replicates. Simulations were used for the cases with known parametric models, whereas a bootstrap based resampling strategy were used for the real data example (cf. Appendix F for the technical details).

**Numerical convergence:** The `R`-package `localgauss`, see Berentsen et al. [2014a], estimates the local Gaussian autocorrelations $\rho_{\boldsymbol{v}}(h)$ and returns them together with an attribute that reveals whether or not the estimation algorithm converged numerically. The $m$-truncated estimates $\hat{f}^m_{\boldsymbol{v}}(\omega)$ inherits the convergence-attributes from the estimates $\{\hat{\rho}_{\boldsymbol{v}}(h)\}_{h=-m}^{m}$, and either "`NC = OK`" or "`NC = FAIL`" will be added to the plot depending on the convergence status. Note that convergence-problems hardly occurs when the computations are based on pseudo-normalized observations.

**Reproducibility and interactive investigations:** All the examples in this paper can be reproduced by the scripts (see Appendix G) that are contained in the `R`-package

`localgaussSpec`. Note that the computations of $\hat{f}^m_{\boldsymbol{v}}(\omega)$ can be performed for a wide range of tuning parameters, which allows an integrated interactive investigation of the results by means of a `shiny`-application.[15]

## 3.2 Estimation aspects for the given parameter configuration

The estimation of $\hat{f}^m_{\boldsymbol{v}}(\omega)$ for a point $\boldsymbol{v} = (v_1, v_2)$ that lies on the diagonal, that is, $v_1 = v_2$, will be based on the estimates of $\hat{\rho}_{\boldsymbol{v}}(h)$ for $h = 1, \ldots, m$, and it is thus of interest to first investigate how these estimates depend on the configuration of the tuning parameters given in section 3.1. This is most easily done in terms of an example, and the pseudo-normalized `dmbp`-data (of length 1974) will be used for this purpose.

First of all, note that the combination of point $\boldsymbol{v}$ and bandwidth $\boldsymbol{b}$ influences how many of the $h$-lagged pairs that effectively contribute to the computation of $\hat{\rho}_{\boldsymbol{v}}(h)$. This is shown in Figure 3 for the pseudo-normalized `dmbp`-data. In the plot of the pseudo-normalized time series (top panel), the three horizontal dashed lines represent the *levels* which corresponds to the coordinates of the three points $\boldsymbol{v}$, whereas the horizontal strips centered at those lines show which observations that lie within a distance of $b = 0.5$ from the respective lines. The three plots at the bottom show the corresponding 1-lagged pairs, each with a *bandwidth-square* (of width $2b$) centered at one of the selected points $\boldsymbol{v}$.

The estimates of $\rho_{\boldsymbol{v}}(1)$ are based on the 1-lagged pairs seen in the lower part of Figure 3, and these and similar estimates for lags up to 200 (based on $\boldsymbol{b} = (0.5, 0.5)$) are shown in Figure 4. An investigation of Figure 4 shows how $\hat{\rho}_{\boldsymbol{v}}(h)$ varies for the three points of interest, and there is a clear distinction between the center and the two tails. Note that the bias-variance balance of the estimates $\hat{\rho}_{\boldsymbol{v}}(h)$ depends on the number of $h$-lagged pairs that effectively contribute during the computation, and it is thus clear that the variance

[15] See Chang et al. [2017] for details about `shiny`.

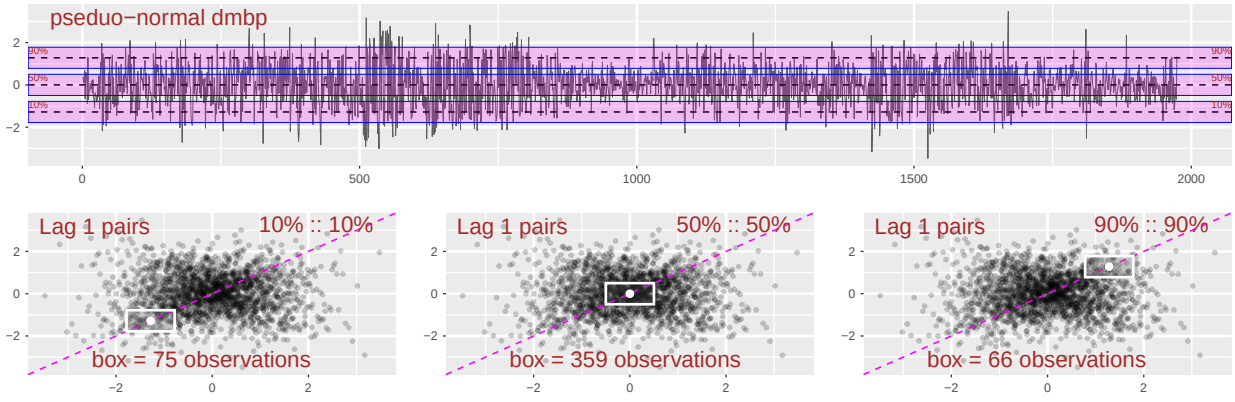


Figure 3: `dmbp` (pseudo-normalized version), *levels* and *bandwidth-bands* (top) and *lag 1 bandwidth-squares* (bottom). Further details in the main text.

will increase for points $\boldsymbol{v}$ that lie farther out in the tails. The selection of which tail-points to investigate must thus take into account the number of available observations for the lags to be included.

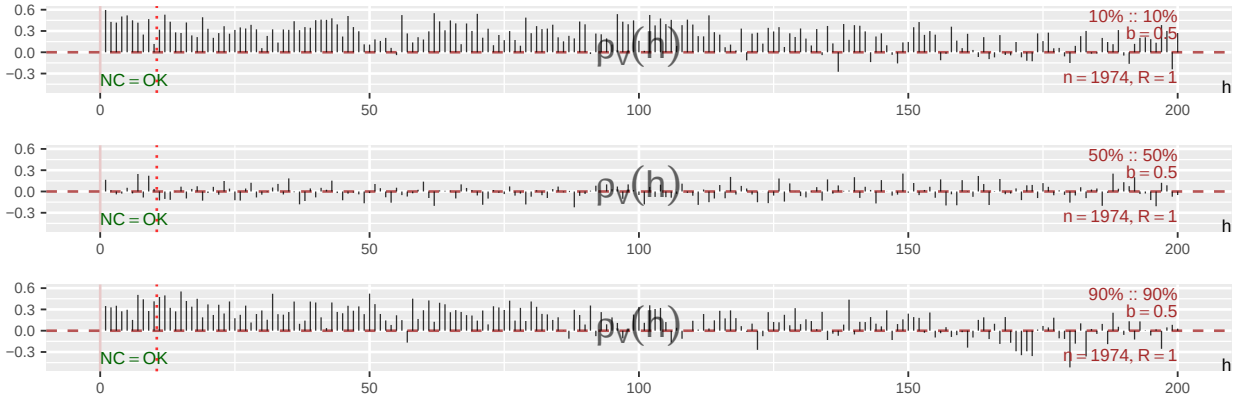


Figure 4: `dmbp`-data, $\hat{\rho}_{\boldsymbol{v}}(h)$ for $h = 1, \ldots, 200$ (for the three points of interest). The estimates for $h = 1, \ldots, 10$ will be used for $\hat{f}^m_{\boldsymbol{v}}(\omega)$, cf. Figure 9.

The $\hat{\rho}_{\boldsymbol{v}}(h)$ tends to fluctuate around 0 at the center, which implies that the corresponding estimated spectral density $\hat{f}_{\boldsymbol{v}}^m(\omega)$ most likely will be rather flat and close to 1. For the two tails, it seems natural to assumme that some long-range dependency must be present, and one might also suspect that there is an asymmetry between the two tails.[16]

Based on the impression from Figure 4, it might be a connection between the global long-range dependence in the `dmbp`-data and the local dependency structure in the tails — but note that the estimates in Figure 4 are based on the pseudo-normalized data, so the information from the marginal distribution is not present here. However, the same kind of behaviour has been observed for pseudo-normalized samples from different GARCH-type models, so the dependency structure of the tails could be a significant contributor to the global long-range dependency seen in time series models like ARCH and GARCH.

## 3.3 Sanity testing the implemented estimation algorithm

The purpose of this section is to check whether or not the implemented estimation algorithm returns reasonable results for some simulated examples. It is only for the Gaussian case that the true value of the local Gaussian spectral densities $f_{\boldsymbol{v}}(\omega)$ are known, and it is thus important to specifically construct an example where heuristic arguments enable the prediction of the anticipated results.

The strategy used to create the plots for the simulated data works as follows: First draw a given number of independent replicates from the specified model, and compute $\hat{f}_{\boldsymbol{v}}^m(\omega)$ and $\hat{f}^m(\omega)$ for each of the replicates. Then extract the mean of these estimates to get estimates of the true values of $f_{\boldsymbol{v}}^m(\omega)$ and $f^m(\omega)$, and select suitable upper and lower percentiles of the estimates to produce an estimate of the pointwise confidence intervals.

[16] A further investigation of this is easy when the `shiny`-application in the `R`-package `localgaussSpec` is used, since it then is possible to immediately switch to an investigation of the corresponding spectra.

Note that the plots have been annotated with the following information: The numerical convergence status NC in the lower left corner; the truncation level $m$ in the upper left corner; the percentiles of the point $\boldsymbol{v}$ of investigation, and the bandwidth $\boldsymbol{b}$ in the upper right corner; the length $n$ and the number of replicates $R$ in the lower right corner.

### 3.3.1 Gaussian white noise

The sanity testing of the implemented estimation algorithm starts with the trivial case. Figure 5 shows the result when the estimation procedure is used on 100 independent samples of length 1974 from a standard normal distribution $\mathrm{N}(0, 1)$. The computations are based on the bandwidth $\boldsymbol{b} = (0.5, 0.5)$, and the points (on the diagonal) corresponds to the 0.1, 0.5 and 0.9 quantiles of the standard normal distribution. The top left panel shows the pseudo-normalized version of the first time series that was sampled from the model, with dashed lines at the levels that corresponds to the above mentioned points. The three other panels contain information about the $m$-truncated ordinary spectral density $f^m(\omega)$ (red part,[17] the same for all the plots) and the $m$-truncated local Gaussian spectral densities $f^m_{\boldsymbol{v}}(\omega)$ for the three points under investigation (blue part).

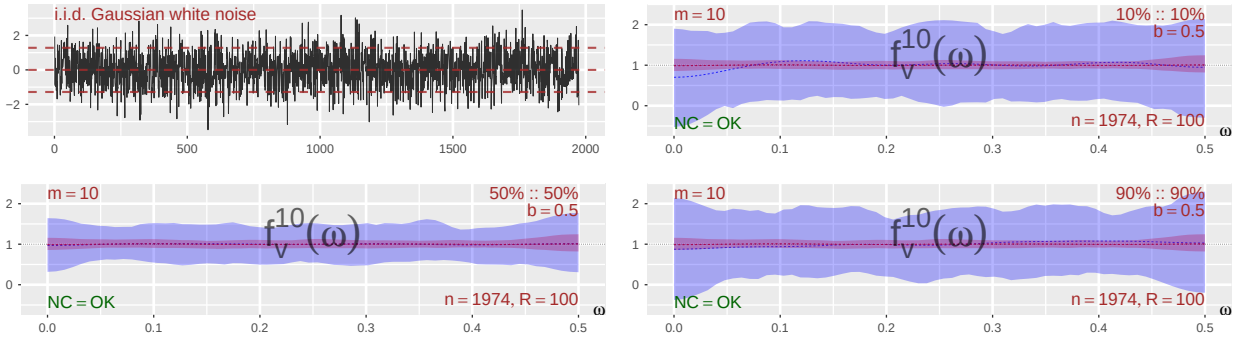


Figure 5: i.i.d. Gaussian white noise, with global and local spectra for three points.

[17] If you have a black and white copy of this paper, then read "red" as "dark" and "blue" as "light".

It can be seen from Figure 5 that the means of the estimates (the dashed lines at the center of the regions) are good estimates of $f^m(\omega)$ and $f_{\boldsymbol{v}}^m(\omega)$, which in this case in fact coincides with $f(\omega)$ and $f_{\boldsymbol{v}}(\omega)$, that is, it is known that the true values are identical to 1 both for the local and global case. Observe that the estimated 90% pointwise confidence intervals are wider for the local Gaussian spectral densities, which is as expected since the bandwidth used in the estimation of the local Gaussian autocorrelations reduces the number of observations that effectively contributes to the estimated values, and thus makes the estimates more prone to small-sample variation. Note also that the pointwise confidence intervals are wider in the tails, which is a natural consequence of the reduced number of points in those regions, cf. the discussion related to Figure 3. The width of these pointwise confidence intervals will decrease when the bandwidth increases, cf. the discussion related to Figure 6.

The estimation procedure gave good estimates of the true values $f(\omega)$ and $f_{\boldsymbol{v}}(\omega)$ in the simple example of Figure 5, but it is important to keep in mind that these plots actually shows estimates of $f^m(\omega)$ and $f_{\boldsymbol{v}}^m(\omega)$. It might be necessary to apply a (much) higher truncation level $m$ before $f^m(\omega)$ and $f_{\boldsymbol{v}}^m(\omega)$ gives decent approximations of the true values $f(\omega)$ and $f_{\boldsymbol{v}}(\omega)$. However, for the task of interest in section 3 it is actually not a problem if the selected truncation level does not give "optimal estimates" of $f(\omega)$ and $f_{\boldsymbol{v}}(\omega)$ — since the detection of nonlinear dependency structures can be seen for a wide range of different truncation levels. The recommended approach is to estimate $\hat{f}_{\boldsymbol{v}}^m(\omega)$ for a range of possible truncation levels $m$, and then check if the shape of the estimates for different truncations share the same properties with regard to the position of any peaks and troughs. The R-package `localgaussSpec` is designed in such a way that this is trivial to do.

### 3.3.2 Some trigonometric examples

Beyond the realm of Gaussian time series, it is not known what the true value for the local Gaussian spectral density actually should be. The sanity of the implemented estimation algorithm will thus be tested by the means of an artificially constructed *local trigonometric* time series, for which it at least can be reasonably argued what the expected outcome should be for some specially designated points $\boldsymbol{v}$ (given a suitable bandwidth $\boldsymbol{b}$). These artificial time series will not satisfy the requirements needed for the asymptotic theory to hold true (as is also the case for standard global spectral analysis), but they can still be used to show how an exploratory tool based on the local Gaussian spectral density can detect local periodic properties that the ordinary spectral density fails to detect.

As a prerequisite (and a reference) for the investigation of the local trigonometric time series, it is prudent to first investigate the result based on independent samples from a time series of the form $Y_t = \cos(2\pi\alpha t + \varphi) + w_t$, where $w_t$ is Gaussian white noise with mean zero and standard deviation $\sigma$, and where it in addition is such that $\alpha$ is fixed for all the replicates, whereas the phase-adjustment $\varphi$ is randomly generated for each individual replicate. A realization with $\alpha = 0.302$ and $\sigma = 0.75$ is shown in Figure 6, where the frequency $\alpha$ has been indicated with a vertical line in order to show that both the local and global approaches in this case have a peak at the expected position. The plots are based on 100 samples of length 1974, and shows 90% pointwise confidence intervals. Some useful remarks can be based on Figure 6, before *the local trigonometric* case is defined and investigated.

All the plots in Figure 6 show the same point (corresponding to the 10% quantile) in the lower tail, but they differ with regard to the bandwidths that have been used. In particular, the upper right plot is based on the bandwidth $\boldsymbol{b} = (.5, .5)$ (the bandwidth used in all the other examples), whereas the two plots at the bottom shows the situation for the

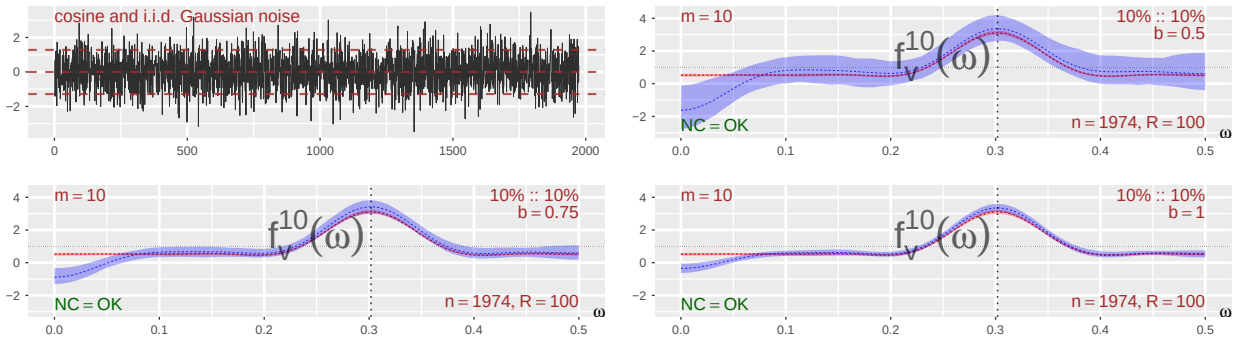


Figure 6: Single cosine and i.i.d. white noise, same point, bandwidths 0.5, 0.75 and 1.

bandwidths $\boldsymbol{b} = (.75, .75)$ and $\boldsymbol{b} = (1, 1)$, respectively at the left and right. In this case, the widths of the pointwise confidence intervals are influenced by the selected bandwidths, but the overall shape is similar and close to the global estimate shown in red. This feature is also present for the other examples that have been investigated.

Note that the cosine is recovered using just a neighbourhood of the 10% quantile. Furthermore, the portion of the local Gaussian spectral density that is negative decreases with increasing bandwidth, which is in accordance with the remark at the end of section 2.1.1. Using the notation from Algorithm 2.1, this can for the estimates of the local Gaussian autocorrelations be stated as $\hat{\rho}_{\boldsymbol{v}}(h|\boldsymbol{b}) \to \hat{\rho}\,(h)$ when $\boldsymbol{b} \to \boldsymbol{\infty}$, which implies that the estimate $\hat{f}^m_{\boldsymbol{v}}(\omega)$ converges towards the global non-negative estimate $\hat{f}^m(\omega)$. It is thus possible to reduce the amount of negative values for the estimates $\hat{f}^m_{\boldsymbol{v}}(\omega)$ by increasing the bandwidth $\boldsymbol{b}$, but keep in mind that it is the limits $\boldsymbol{b} \to \boldsymbol{0}^+$ and $m \to \infty$ that should be taken in order to actually estimate the local Gaussian spectral density $f_{\boldsymbol{v}}(\omega)$.

The truncation level used in Figure 6 is rather low, that is, $m = 10$, but it can be seen that the peak is observed at the correct frequency. The peak will grow taller and narrower when a higher truncation level is used, but it will stay at the same frequency. This indicates that these plots (even for low truncation values) can detect properties of the

underlying structure. Again, this feature is shared with the other examples that have been investigated.

The local Gaussian spectral densities in Figure 6 goes below zero for low frequencies, a feature that is not entirely unexpected as $\{\rho_{\boldsymbol{v}}(h)\}_{h\in\mathbb{Z}}$, the collection of local Gaussian autocorrelations, may not be a non-negative definite function. In fact, based on the observation that the estimates of $\hat{f}^m_{\boldsymbol{v}}(\omega)$ have peaks that are taller and wider than those of $\hat{f}^m(\omega)$, it is as expected that these estimates might need to have negative values somewhere. The reason for this is that all the spectral densities (global, local and $m$-truncated) by construction necessarily must integrate to one over the interval $(-\frac{1}{2}, \frac{1}{2}]$. The higher and wider peaks of the estimates for $\hat{f}^m_{\boldsymbol{v}}(\omega)$ thus requires that it has to lie below the estimates of $\hat{f}^m(\omega)$ in some other region, and if necessary it must attain negative values somewhere. The interesting details in the plots are thus the position of the peaks of $\hat{f}^m_{\boldsymbol{v}}(\omega)$, and regions with negative values should not in general be considered a too troublesome feature.

Note that, under certain circumstances, $\hat{f}^m_{\boldsymbol{v}}(\omega)$ might contain spurious artefacts when it is computed for time series having a non-flat ordinary spectrum, c.f. Appendix G.4.4 for a discussion related to a case based on a deterministic function with small noise.

**The local trigonometric case:** The key idea in this example is that an artificial time series $\{Y_t\}_{t\in\mathbb{Z}}$ can be constructed by the following scheme:

1. Select $r$ time series $\{C_i(t)\}_{i=1}^r$.
2. Select a random variable $I$ with values in the set $\{1, \ldots, r\}$, and use this to sample a collection of indices $\{I_t\}_{t\in\mathbb{Z}}$ (that is, for each $t$ an independent realization of $I$ is taken). Let $p_i := \mathrm{P}(I_i = i)$ denote the probabilities for the different outcomes.
3. Define $Y_t$ by means of the equation

$$Y_t := \sum_{i=1}^{r} \mathbb{1}\{I_t = i\} \cdot C_i(t). \tag{28}$$

The indicator function $\mathbb{1}\{\cdot\}$ ensures that only one of the $C_i(t)$ contribute for a given value $t$, that is, it is also possible to write $Y_t = C_{I_t}(t)$.

The *local trigonometric* time series (needed for the sanity testing of the implemented estimation algorithm) are constructed by selecting $r$ cosine-functions that oscillates around different horizontal base-lines $L_i$, that is,

$$C_i(t) = L_i + A_i(t) \cdot \cos\left(2\pi\alpha_i t + \varphi_i\right), \qquad i = 1, \ldots, r, \tag{29}$$

where $\alpha_i$ and $\varphi_i$ respectively represent the frequency and phase-adjustment occurring in the cosine-function, and where the amplitudes $A_i(t)$ are uniformly distributed in some interval $[a_i, b_i]$. Note that it is assumed that the phases $\varphi_i$ are uniformly drawn (one time for each realization) from the interval between 0 and $2\pi$, and it is moreover also assumed that the stochastic processes $\varphi_i$, $A_i(t)$ and $I_t$ are independent of each other.

The autocorrelation $\rho(h)$ of the time series $\{Y_t\}_{t\in\mathbb{Z}}$, with $C_i(t)$ as given in Equation (29), has been computed in the Supplementary Material, cf. Equation (G.5) in Appendix G.4. For the purpose of the present section, it is sufficient to know that it is possible to find parameter-configurations for which the global spectrum is rather flat (at least when truncated at $m = 10$), which implies that it cannot detect the frequencies $\alpha_i$ of the underlying structure.

Strictly speaking, neither $f(\omega)$ nor $f_{\boldsymbol{v}}(\omega)$ are well defined for the *local trigonometric* times series, but this is not important since it still is possible to predict (cf. Appendix G.4 for details) that the $m$-truncated estimates $\hat{f}^m_{\boldsymbol{v}}(\omega)$ for some points $\boldsymbol{v}$ should resemble Figure 6 — and this can be used, cf. Figure 7, to test the sanity of the implemented estimation algorithm.

The explicit expression for the *local trigonometric* example studied in Figure 7 is given by $r = 4$ components $C_i(t)$ of the form given in Equation (29), where the probabilities $p_i$ are

given by $(0.05, 1/3-0.05, 1/3, 1/3)$, the frequencies $\alpha_i$ are given by $(0.267, 0.091, 0.431, 0.270)$, the base-lines $L_i$ are given by the values $(-2, -1, 0, 1)$, and the lower and upper ranges for the uniforms sampling of the amplitudes $A_i(t)$ are respectively given by $(0.5, 0.2, 0.2, 0.5)$ and $(1.0, 0.5, 0.3, 0.6)$. Note that $L_i$ and $A_i(t)$ should be selected in order to give a minimal amount of overlap between the different components, cf. Appendix G.4 for further details.

Figure 7 shows $\hat{f}^m(\omega)$ and $\hat{f}^m_{\boldsymbol{v}}(\omega)$ for the *local trigonometric* example. The ordinary spectrum does not detect the frequencies $\alpha_i$ (indicated by vertical lines), whereas the local Gaussian spectra does have clear peaks at the frequencies from respectively $C_2(t)$, $C_3(t)$ and $C_4(t)$. Moreover, a comparison with Figure 6 shows that $\hat{f}^m_{\boldsymbol{v}}(\omega)$ indeed does look like predicted, which verifies the sanity of the implemented estimation algorithm.

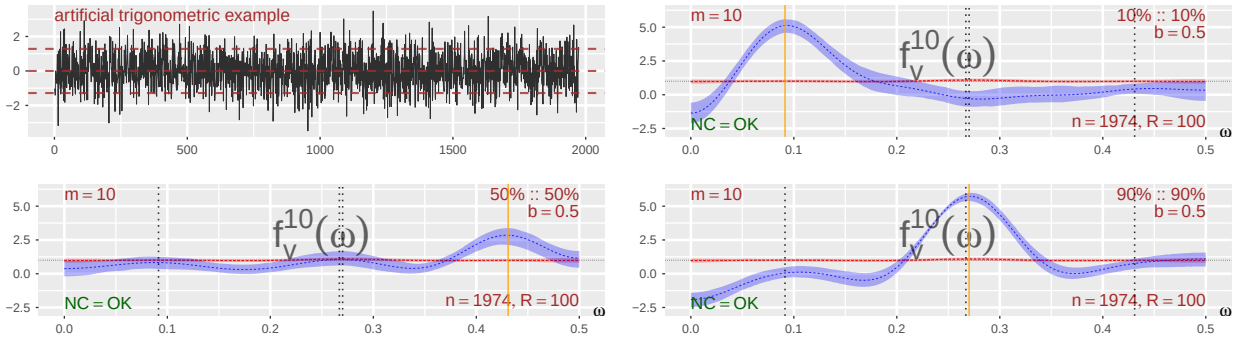


Figure 7: Artifical example, *local trigonometric components*. Global and local spectra for the three points $\boldsymbol{v}$ on the diagonal, that is, lower tail, center and upper tail.

The selected percentiles $\{p_i\}_{i=1}^4$ implies that observations from the $C_3(t)$ component after pseudo-normalization should lie between $\Phi^{-1}(1/3) = -0.43$ and $\Phi^{-1}(2/3) = 0.43$. The estimation of $f^m_{\boldsymbol{v}}(\omega)$ is based on the bandwidth $\boldsymbol{b} = (0.5, 0.5)$, which implies that the estimate at the center will be "contaminated" by observations from the neighbouring components — and this explains the lower amplitude seen for this point.

The three points $\boldsymbol{v}$ in Figure 7 correspond roughly to the base-lines $L_2, L_3$ and $L_4$, and the corresponding frequencies $\alpha_2, \alpha_3$ and $\alpha_4$ are here detected by $\hat{f}^m_{\boldsymbol{v}}(\omega)$. But what about the base-line $L_1$ and the $\alpha_1$-frequency?

The low probability at which the $C_1(t)$ component is selected implies that the point $\boldsymbol{v}$ corresponding to the base-line $L_1$ must lie far out in the lower tail, and for the present sample size (of $n = 1974$) the scarcity of observations in this region implies that it is not possible to obtain decent estimates of the required local Gaussian autocorrelations $\rho_{\boldsymbol{v}}(h)$. A countermeasure to this problem would be to use a larger bandwidth $\boldsymbol{b}$, but the result would then be "contaminated" by the observations from the $C_2(t)$ component — and the peak of $\hat{f}^m_{\boldsymbol{v}}(\omega)$ would then be at the frequency $\alpha_2$ instead of $\alpha_1$. This implies that misleading results can occur when the bandwidth $\boldsymbol{b}$ is to large.

However, note that for a large enough sample it is possible to detect the frequency $\alpha_1$ that belongs to the $C_1(t)$-component, cf. Appendix G.4.3 for further details.

The $C_1(t)$ component was included in this example in order to emphasize that extra care is needed when investigating the outer tails of a sample. This of course begs the question: For a given sample $\{Y_t\}_{t=1}^n$, how can an investigator figure out whether or not the estimate of $f^m_{\boldsymbol{v}}(\omega)$, for a given combination of point $\boldsymbol{v}$ and bandwidth $\boldsymbol{b}$, seems trustworthy or not? Another important question for an investigator is to decide if some points $\boldsymbol{v}$ might be more interesting than others. Both of these questions can be investigated by means of the two plots seen in Figure 8, which (for a single sample from the aforementioned *local trigonometric* construction) investigates the $m = 10$ truncated local Gaussian spectra $f^m_{\boldsymbol{v}}(\omega)$ for points along the diagonal. Note that the points $\boldsymbol{v}$ are represented by their respective percentiles, and the range goes from the 5% percentile to the 95% percentile.

The upper part of Figure 8 is a heatmap-plot for $\hat{f}^m_{\boldsymbol{v}}(\omega)$ (inspired by plots in Li [2019]), which in this case is based on one sample of length $n = 1974$. The contour-lines in this

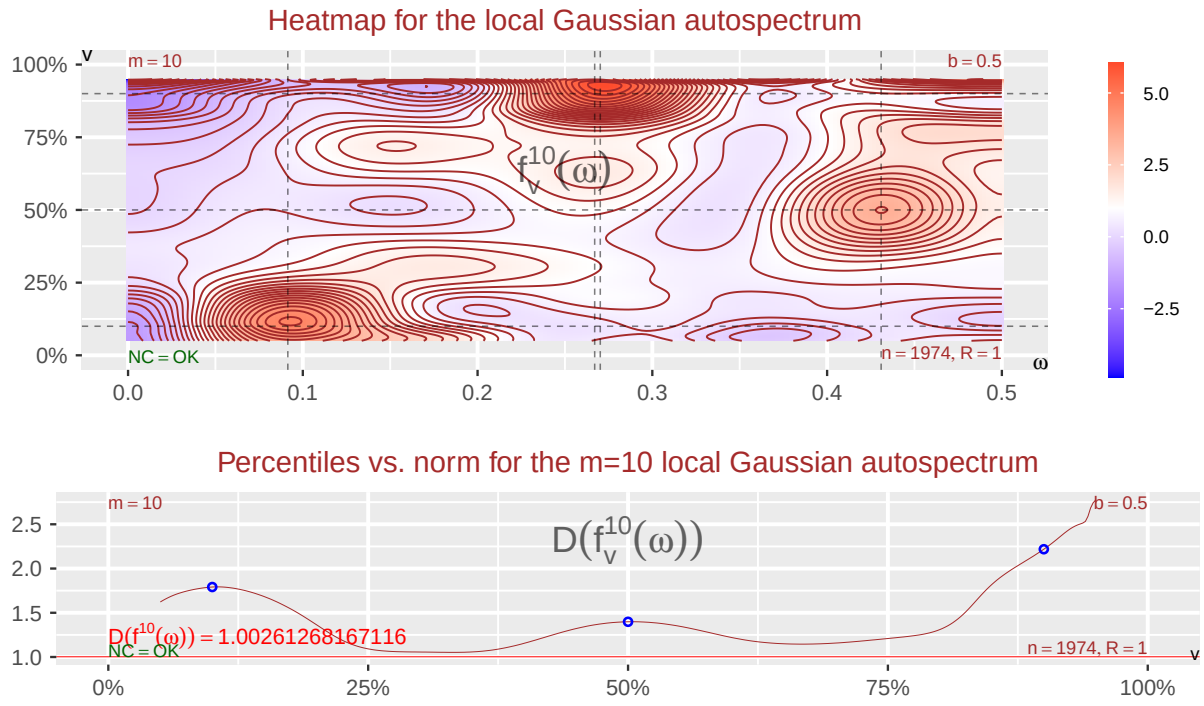


Figure 8: Heatmap-plot with corresponding distance-plot, based on the local trigonometric case, showing how $\hat{f}_{\boldsymbol{v}}^{10}(\omega)$ varies with the percentiles for the diagonal-points $\boldsymbol{v}$. The percentiles and frequencies used in Figure 7 have been indicated with lines/points.

plot clearly reveals that the highest peaks occur approximately at the points investigated in Figure 7. In fact, looking at the heatmap, the peak at the 90% percentile of Figure 8, may have its maximum closer to the 95% percentile, but one has to be a little careful here since the estimates of $\rho_{\boldsymbol{v}}(h)$ might degenerate towards $+1$ (or $-1$) in the outer part of the tail.

The lower part of Figure 8 shows the corresponding distance-plot $D(f_{\boldsymbol{v}}^m(\omega))$, where the norms of the $m$-truncated spectra (realized as elements of the complex Hilbert space of Fourier series, cf. Appendix D.1 for details) are plotted against the diagonal points. Note that distance-based plots do not contain any information about the frequencies, and completely different spectral densities can have the same distance-value. It is thus important to always combine a distance-based plot with a plot that reveals the frequency-component.

The horizontal line at the bottom of the distance-plot gives the norm of the ordinary spectrum, and it can be seen that this line is very close to the white-noise value which is 1. It is interesting and reassuring that it picks up the peaks at the 10% and 50% percentiles. It does however not indicate a peak close to the 95% percentile, but this is also the least clear peak of the heatmap.

This discussion shows that it is important to include a wide range of points when performing an investigation based on local Gaussian spectral densities, since it is necessary to check how $\hat{f}_{\boldsymbol{v}}^m(\omega)$ changes as the diagonal point $\boldsymbol{v}$ varies from the lower tail to the upper tail. The `R`-package `localgaussSpec` is designed for such investigations, and it includes an interactive interface that can switch between different visualizations. Note that `localgaussSpec` also can deal with points $\boldsymbol{v}$ that lies outside of the diagonal, and it can in addition also digest multivariate time series.

## 3.4 Real data and a fitted GARCH-type model

The local Gaussian machinery will now be used on the `dmbp`-data. It will here be seen that local properties of the nonlinear dependency structure indeed can be obtained by comparing $\hat{f}^m(\omega)$ and $\hat{f}_{\boldsymbol{v}}^m(\omega)$, and this works even for low values of the truncation level $m$.

Another topic that it is natural to consider is the comparison of $\hat{f}_{\boldsymbol{v}}^m(\omega)$ based on the data and $\hat{f}_{\boldsymbol{v}}^m(\omega)$ based on simulations from a model fitted to the data — and this will in

particular be investigated for a GARCH-type model that was fitted to the `dmbp`-data by the `R`-package `rugarch`, Ghalanos [2020].

### 3.4.1 The real data example

The `dmbp`-data (length 1974), whose original and pseudo-normalized versions can be seen in Figure 2, will now be investigated by the $m$-truncated local Gaussian spectral densities $f_{\boldsymbol{v}}^m(\omega)$. These estimates will be based on the bandwidth $\boldsymbol{b} = (0.5, 0.5)$, and they will be computed for the three diagonal points corresponding to the 10%, 50% and 90% percentiles of the standard normal distribution. The estimated local Gaussian autocorrelations $\hat{\rho}_{\boldsymbol{v}}(h)$ that is used in the computation of $\hat{f}_{\boldsymbol{v}}^m(\omega)$ can be seen in Figure 4, and the estimated values of $\hat{f}^m(\omega)$ and $\hat{f}_{\boldsymbol{v}}^m(\omega)$ (for the $m = 10$ case) are shown as the red and blue solid lines[18] in Figure 9. The pointwise confidence intervals are based on the resampling strategy discussed at the end of this section.

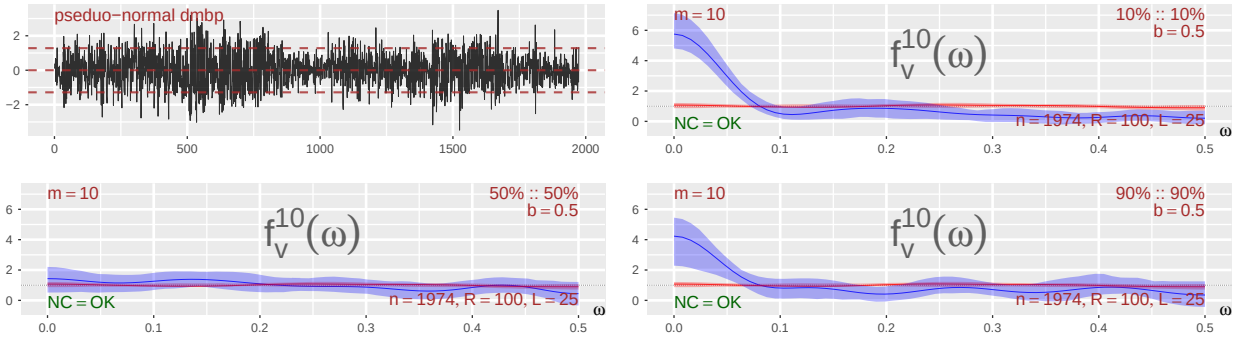


Figure 9: `dmbp`-data, bootstrapped based confidence intervals. Global and local spectra for the three diagonal points.

The global spectrum $\hat{f}^m(\omega)$ is flat, which is in agreement with the knowledge that the `dmbp`-data resembles white noise. The local Gaussian spectrum $\hat{f}_{\boldsymbol{v}}^m(\omega)$ at the center is also

[18] Solid lines are always used by the `R`-package `localgaussSpec` when $\hat{f}_{\boldsymbol{v}}^m(\omega)$ is based on real data.

rather flat, which is no surprise given the values $\hat{\rho}_{\boldsymbol{v}}(h)$ seen in the middle panel of Figure 4. The estimates $\hat{f}^m_{\boldsymbol{v}}(\omega)$ in the tails are obviously not flat, and the clear peaks at the frequency $\omega = 0$ are again in agreement with the corresponding values $\hat{\rho}_{\boldsymbol{v}}(h)$ from Figure 4.

The difference between the (solid lines in the) lower and upper tail could indicate the presence of an asymmetry, that is, the peak are more prominent for the lower tail. It would be premature to draw a firm conclusion regarding asymmetry based one a single plot using the low truncation level $m = 10$, but the asymmetry can also be seen for higher truncation levels (investigated up to $m = 200$), with an increasing difference between the height of these peaks. Such an asymmetry, with a higher peak at the lower tail, would be in agreement with the asymmetry between a *bear market* (going down) and a *bull market* (going up).

A comparison solely based on the solid lines in Figure 9 is not sufficient, since an observed difference could be due to the variability of the estimator used to find $\hat{f}^m_{\boldsymbol{v}}(\omega)$. It is thus necessary to decide on a reasonable resampling strategy (described below) that can provide pointwise confidence intervals like those shown in Figure 9. Based on the pointwise confidence intervals, it is clear that the truncated local and global spectra indeed do show that the `dmbp`-data contains local non-linear dependency structures in the tails. Note that the width of the pointwise confidence interval is a function of the frequency, cf. Appendix A.3, and this can in some cases give it a wide "trumpet shape" near $\omega = 0$, as seen in the lower and upper tails in Figure 9 (and which is even more prominent in Figure 11)

The pointwise confidence intervals in Figure 9 requires a resampling strategy that takes into account that the local Gaussian autocorrelations $\rho_{\boldsymbol{v}}(1), \ldots, \rho_{\boldsymbol{v}}(m)$ are estimated by a local likelihood approach. The asymptotic properties of these estimates were developed in the present paper using the procedure from Klimko and Nelson [1978], cf. Appendix B.1.

The block bootstrap can be used for a variety of estimators, and it can in particular, cf. Künsch [1989, Example 2.4, p. 1219-20], be applied for estimators based on the Klimko-Nelson procedure. The block bootstrap was thus used as the resampling strategy in an earlier draft of this paper, and the results were similar to Figure 9 when a block length of $L = 100$ was used. The selected block length $L$ seemed reasonable based on the $\hat{\rho}_{\boldsymbol{v}}(h)$-values seen in Figure 4. See Appendix F.6 for further details.

Some comments related to the block bootstrap were received during the review-process, and those motivated the investigation presented in Appendix F, which lead to the adjusted resampling strategy given in Algorithm F.4. The adjusted resampling method uses a two step procedure, where the first step uses the block bootstrap on the *indices of the observations*, and the next step uses those resampled indices to identify the $h$-lagged pairs $\big(Y_{t+h}, Y_t\big)$ that should be used when estimating $\rho_{\boldsymbol{v}}(h)$ for the resampled data.

The adjusted resampling approach reduce the edge-effect noise that occurs when the components of a resampled pair belong to different blocks, and this implies that it works well with lower block lengths than those needed for the ordinary block bootstrap. A sensitivity analysis related to the selection of the block length $L$ is presented in Appendix F.5.

### 3.4.2 A heatmap/distance plot for the `dmbp`-data

It is of interest to know how $\hat{f}^m_{\boldsymbol{v}}(\omega)$ behaves for other diagonal points, and this can be seen in Figure 10 which is constructed in the same manner as Figure 8. Keep in mind that these plots are based on pseudo-normalized data, that is, the information in the marginal distribution is not present, and Figure 10 thus primarily reveals information about the copula-structure of the time series under investigation, cf. Appendix D.2 for further details.

Figure 10 supports the impression that there is an asymmetry between the lower tail and the upper tail, and it can also be seen that the local dependency structure is weak near the center. Note that these plots go from the 5% to 95% percentile, in order to show that it

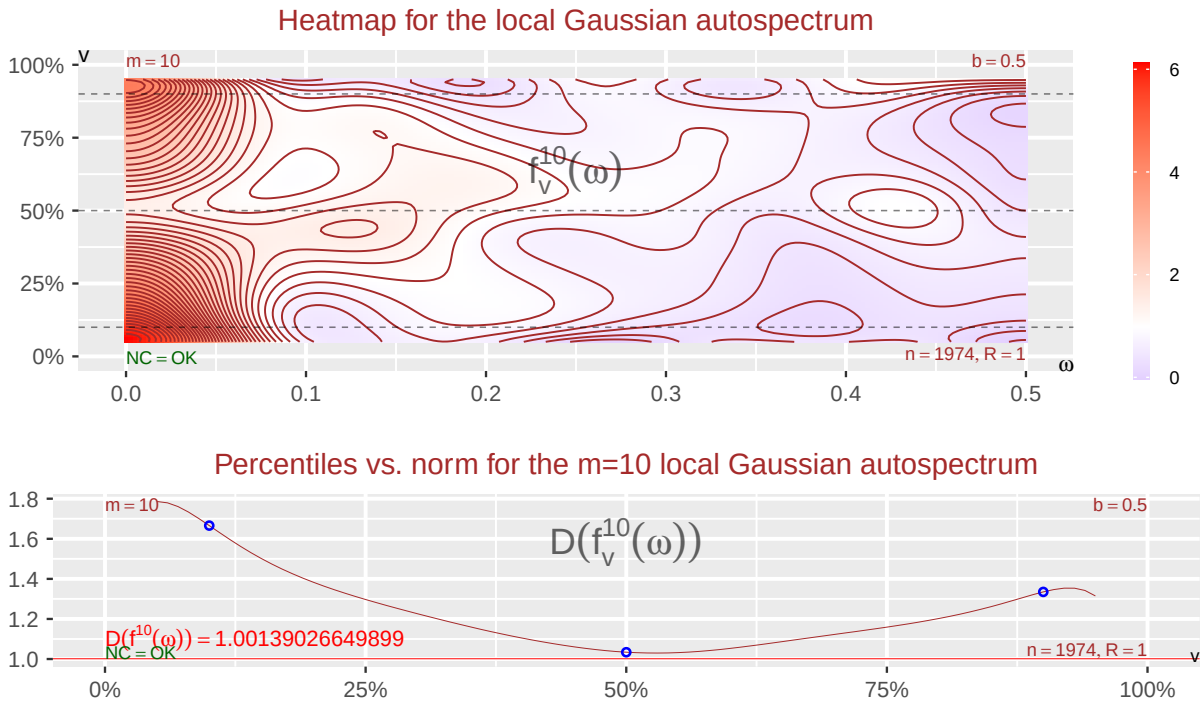


Figure 10: Heatmap and corresponding distance-based plots based on the `dmbp-data`, showing how $\hat{f}^{10}_{\boldsymbol{v}}(\omega)$ varies with the percentiles for the diagonal-points $\boldsymbol{v}$. The percentiles used in Figure 9, that is, 10%, 50% and 90%, have been highlighted with lines/points.

might be perilous to go too far out in the tail for the present sample size ($n = 1974$). This is discussed in more detail in Appendix D.2, where heatmap based plots of the estimated underlying local Gaussian autocorrelations can be found, cf. Figures D.2 to D.4.

### 3.4.3 A GARCH-type model

This section will consider an *asymmetric power ARCH-model* (apARCH) of order $(2, 3)$, with parameters based on a fitting to the `dmbp`-data.[19] Technical details about this model, and comments regarding the script needed for the reproduction of this example, can be found in Appendix G.3 in the Supplementary Material.

For a comparison with the results based on the `dmbp`-data, it is natural to consider $R = 100$ samples of length $n = 1974$ from the fitted apARCH$(2, 3)$ model — and the estimates of $f^m_{\boldsymbol{v}}(\omega)$ should be computed for the same points $\boldsymbol{v}$ and with the same tuning parameters $\boldsymbol{b}$ and $m$. The result from such an investigation can be seen in Figure 11.

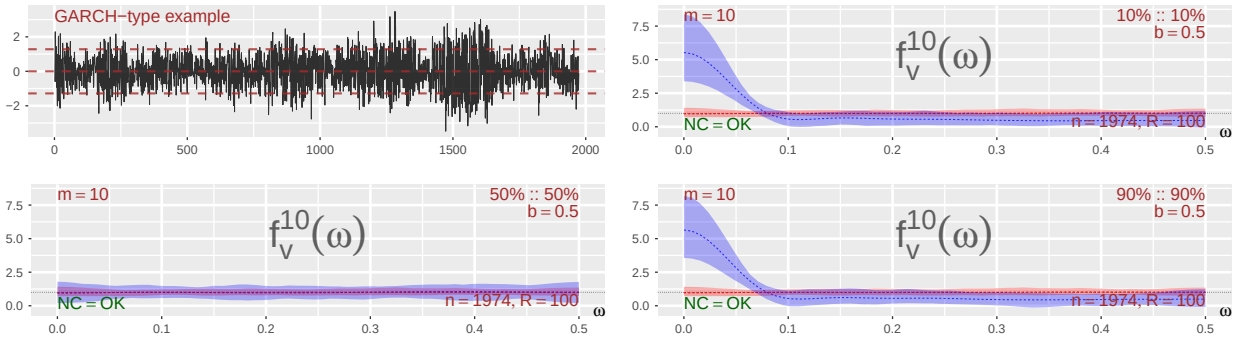


Figure 11: GARCH-type model, based on `dmbp`. Global and local spectra for three points.

It is clear from Figure 11 that the estimate of the $m$-truncated global spectrum is flat, and this is in agreement with the knowledge that $f(\omega) = 1$ for a GARCH-type model (since $\rho(h) = 0$ when $h \neq 0$). It can also be seen that the esimates $\hat{f}^m_{\boldsymbol{v}}(\omega)$ based on the fitted model have the same overall structure as those in Figure 9. In particular, there is a flat spectrum at the center, and the tails show the presence of nonlinear structures with peaks

[19] The `R`-package `rugarch`, Ghalanos [2020] was used to find the parameters of a multitude of GARCH-type models, and the asymmetric power ARCH model with the best fit was then selected.

at $\omega = 0$. Figure 11 does however not pick up the apparent and intuitively reasonable asymmetry seen in the solid lines in Figure 9, which also are supported by the plots in Figure 10.

### 3.4.4 Local testing of fitted models

A comparison of plots like those in Figures 9 and 11 can be used to perform a "local sanity check" of whether or not the dependency structure of the fitted model properly matches the dependency structure of the data — and it is also possible to perform "local comparisons" of different models that have been fitted to the same data. The interested reader can find similar local investigations of data and fitted models in for example Birr et al. [2019]; Li [2019].

Note that it for such comparisons also is of interest to include points $\boldsymbol{v}$ outside the diagonal. The plots needed for off-diagonal points must take into account that $\hat{f}^m_{\boldsymbol{v}}(\omega)$ will be complex-valued outside the diagonal, but this has already been taken care of in the R-package `localgaussSpec`, where the implemented solution simply mimics the co-spectra, quadrature-spectra, phase-spectra and amplitude spectra that is used for the ordinary complex-valued cross-spectra.

An alternative strategy to the comparison of two sets of plots, like those in Figures 9 and 11, is to superimpose the $\hat{f}^m_{\boldsymbol{v}}(\omega)$ from the `dmbp`-data on the top of the corresponding plots based on the fitted model. A plot based on this superposition principle (inspired by a similar plot from Birr et al. [2019]) is given in Figure F.1 in the Supplementary Material, cf. Appendix F.2. Note that this plot also contains visualizations of complex-valued spectra.

# 4 Conclusion

The local Gaussian spectral density $f_{\boldsymbol{v}}(\omega)$ has in this paper been introduced as a new tool for the study of nonlinear time-series. The examples show that even for low truncation levels it

is possible to detect nonlinear periodicities missed by the ordinary spectral density. Further, one can detect the presence of general nonlinear dependency structures by a comparison of the $m$-truncated versions of the ordinary spectrum and the local Gaussian spectra.

The $m$-truncated spectra $f_{\boldsymbol{v}}^m(\omega)$ can also be of interest with regard to *local comparisons* of models fitted to a given sample, as discussed at the end of section 3.4.

The `R`-package `localgaussSpec` can estimate $f_{\boldsymbol{v}}^m(\omega)$ for a large number of combinations of points $\boldsymbol{v}$, truncation levels $m$, and block lengths $\boldsymbol{b}$ — and it does also have an integrated `shiny`-application that enables an easy interactive investigation of the results. The Supplementary Material contains a sensitivity analysis that shows how $\hat{f}_{\boldsymbol{v}}^m(\omega)$ reacts to adjustments of $\boldsymbol{v}$, $m$ and $\boldsymbol{b}$ — and it is there also seen that adjustments of the block length $L$, within wide intervals, have a minimal impact on the pointwise confidence intervals.

## Acknowledgements

The authors are most grateful for the valuable comments and suggestions from the referees and the associate editor.

## Supplementary Material

The online Supplementary Material contains the appendices. The scripts needed for the reproduction of the examples in this paper is contained in the `R`-package `localgaussSpec`, cf. Appendix G for further details.

# References


Berentsen, G.D., Cao, R., Francisco-Fernández, M., Tjøstheim, D., 2017. Some Properties of Local Gaussian Correlation and Other Nonlinear Dependence Measures. Journal of Time Series Analysis 38, 352–380. URL: http://dx.doi.org/10.1111/jtsa.12183, doi:10.1111/jtsa.12183.

Berentsen, G.D., Kleppe, T.S., Tjøstheim, D.B., 2014a. Introducing `localgauss`, an R Package for Estimating and Visualizing Local Gaussian Correlation. j-J-STAT-SOFT 56. URL: http://www.jstatsoft.org/v56/i12, doi:10.18637/jss.v056.i12.

Berentsen, G.D., Tjøstheim, D., 2014. Recognizing and visualizing departures from independence in bivariate data using local Gaussian correlation. Statistics and Computing 24, 785–801. URL: http://dx.doi.org/10.1007/s11222-013-9402-8, doi:10.1007/s11222-013-9402-8.

Berentsen, G.D., Tjøstheim, D., Nordbø, T., 2014b. Recognizing and visualizing copulas: An approach using local Gaussian approximation. Insurance: Mathematics and Economics 57, 90 – 103. URL: http://www.sciencedirect.com/science/article/pii/S0167668714000432, doi:10.1016/j.insmatheco.2014.04.005.

Birr, S., Kley, T., Volgushev, S., 2019. Model assessment for time series dynamics using copula spectral densities: A graphical tool. Journal of Multivariate Analysis 172, 122 – 146. URL: http://www.sciencedirect.com/science/article/pii/S0047259X18301842, doi:10.1016/j.jmva.2019.03.003. Dependence Models.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307 – 327. URL: http://www.sciencedirect.com/science/article/pii/0304407686900631, doi:10.1016/0304-4076(86)90063-1.

Bollerslev, T., Ghysels, E., 1996. Periodic Autoregressive Conditional Heteroscedasticity. Journal of Business & Economic Statistics 14, 139–151. URL: http://amstat.tandfonline.com/doi/abs/10.1080/07350015.1996.10524640, doi:10.1080/07350015.1996.10524640, arXiv:http://amstat.tandfonline.com/doi/pdf/10.1080/07350015.1996.10524640.

Brillinger, D.R. (Ed.), 1984. The collected works of John W. Tukey. Volume I. Time series: 1949–1964. Wadsworth Statistics/Probability Series, Wadsworth, Pacific Grove, CA, USA. With introductory material by William S. Cleveland and Frederick Mosteller.

Brillinger, D.R., 1991. Some history of the study of higher-order moments and spectra. Statistica Sinica 1, 24J. URL: http://www3.stat.sinica.edu.tw/statistica/j1n2/j1n23/..\j1n210\j1n210.htm.

Carrasco, M., Chen, X., 2002. Mixing and moment properties of various garch and stochastic volatility models. Econometric Theory 18, 17–39. doi:10.1017/S0266466602181023.

Chang, W., Cheng, J., Allaire, J., Xie, Y., McPherson, J., 2017. shiny: Web Application Framework for R. URL: https://CRAN.R-project.org/package=shiny. r package version 1.0.3.

Davis, R.A., Mikosch, T., 2009. The extremogram: A correlogram for extreme events. Bernoulli 15, 977–1009. URL: https://doi.org/10.3150/09-BEJ213, doi:10.3150/09-BEJ213.

Ghalanos, A., 2020. rugarch: Univariate GARCH models. URL: https://cran.r-project.org/package=rugarch. r package version 1.4-2.

Hagemann, A., 2011. Robust Spectral Analysis. Available at SSRN 1956581 URL: https://ssrn.com/abstract=1956581, doi:10.2139/ssrn.1956581.

Han, H., Linton, O., Oka, T., Whang, Y.J., 2016. The cross-quantilogram: Measuring quantile dependence and testing directional predictability between time series. Journal of Econometrics 193, 251 – 270. URL: http://www.sciencedirect.com/science/article/pii/S0304407616300458, doi:10.1016/j.jeconom.2016.03.001.

Hjort, N.L., Jones, M.C., 1996. Locally parametric nonparametric density estimation. Ann. Statist. 24, 1619–1647. URL: http://dx.doi.org/10.1214/aos/1032298288, doi:10.1214/aos/1032298288.

Hong, Y., 1999. Hypothesis Testing in Time Series via the Empirical Characteristic Function: A Generalized Spectral Density Approach. Journal of the American Statistical Association 94, 1201–1220. URL: http://tandfonline.com/doi/abs/10.1080/01621459.1999.10473874, doi:10.1080/01621459.1999.10473874, arXiv:http://tandfonline.com/doi/pdf/10.1080/01621459.1999.10473874.

Hong, Y., 2000. Generalized spectral tests for serial dependence. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 62, 557–574. URL: http://onlinelibrary.wiley.com/doi/10.1111/1467-9868.00250/abstract, doi:10.1111/1467-9868.00250.

Jordanger, L.A., Tjøstheim, D., 2017. Nonlinear cross-spectrum analysis via the local gaussian correlation. arXiv preprint arXiv:1708.02495 URL: https://arxiv.org/abs/1708.02495, arXiv:https://arxiv.org/abs/1708.02495.

Klimko, L.A., Nelson, P.I., 1978. On Conditional Least Squares Estimation for Stochastic Processes. Ann. Statist. 6, 629–642. URL: http://dx.doi.org/10.1214/aos/1176344207, doi:10.1214/aos/1176344207.

Klüppelberg, C., Mikosch, T., 1994. Some Limit Theory for the Self-Normalised Periodogram of Stable Processes. Scandinavian Journal of Statistics 21, 485–491. URL: http://www.jstor.org/stable/4616332.

Künsch, H.R., 1989. The Jackknife and the Bootstrap for General Stationary Observations. The Annals of Statistics 17, 1217–1241. URL: http://www.jstor.org/stable/2241719, doi:10.1214/aos/1176347265.

Lacal, V., Tjøstheim, D., 2017. Local Gaussian Autocorrelation and Tests for Serial Independence. Journal of Time Series Analysis 38, 51–71. URL: http://dx.doi.org/10.1111/jtsa.12195, doi:10.1111/jtsa.12195. 10.1111/jtsa.12195.

Lacal, V., Tjøstheim, D., 2018. Estimating and Testing Nonlinear Local Dependence Between Two Time Series. Journal of Business & Economic Statistics 0, 1–13. URL: https://doi.org/10.1080/07350015.2017.1407777, doi:10.1080/07350015.2017.1407777, arXiv:https://doi.org/10.1080/07350015.2017.1407777.

Li, H., Zhong, W., Park, S.Y., 2016. Generalized cross-spectral test for nonlinear Granger causality with applications to money–output and price–volume relations. Economic Modelling 52, Part B, 661 – 671. URL: http://www.sciencedirect.com/science/article/pii/S0264999315002916, doi:10.1016/j.econmod.2015.09.037.

Li, T.H., 2008. Laplace Periodogram for Time Series Analysis. Journal of the American Statistical Association 103, 757–768. URL: http:

//dx.doi.org/10.1198/016214508000000265, doi:10.1198/016214508000000265, arXiv:http://dx.doi.org/10.1198/016214508000000265.

Li, T.H., 2010a. A Nonlinear Method for Robust Spectral Analysis. IEEE Transactions on Signal Processing 58, 2466–2474. URL: http://ieeexplore.ieee.org/abstract/document/5406102/, doi:10.1109/TSP.2010.2042479.

Li, T.H., 2010b. Robust coherence analysis in the frequency domain, in: Signal Processing Conference, 2010 18th European, IEEE. pp. 368–371. URL: http://ieeexplore.ieee.org/abstract/document/7096642/.

Li, T.H., 2010c. A robust periodogram for high-resolution spectral analysis. Signal Processing 90, 2133 – 2140. URL: http://www.sciencedirect.com/science/article/pii/S0165168410000137, doi:10.1016/j.sigpro.2010.01.012.

Li, T.H., 2012a. Detection and estimation of hidden periodicity in asymmetric noise by using quantile periodogram, in: 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3969–3972. URL: http://ieeexplore.ieee.org/abstract/document/6288787/, doi:10.1109/ICASSP.2012.6288787.

Li, T.H., 2012b. On robust spectral analysis by least absolute deviations. Journal of Time Series Analysis 33, 298–303. URL: http://dx.doi.org/10.1111/j.1467-9892.2011.00760.x, doi:10.1111/j.1467-9892.2011.00760.x.

Li, T.H., 2012c. Quantile Periodograms. Journal of the American Statistical Association 107, 765–776. URL: http://dx.doi.org/10.1080/01621459.2012.682815, doi:10.1080/01621459.2012.682815.

Li, T.H., 2014. Quantile Periodogram and Time-Dependent Variance. Journal of Time Series Analysis 35, 322–340. URL: http://dx.doi.org/10.1111/jtsa.12065, doi:10.1111/jtsa.12065.

Li, T.H., 2019. Quantile-Frequency Analysis and Spectral Divergence Metrics for Diagnostic Checks of Time Series With Nonlinear Dynamics. Papers. arXiv.org. URL: https://EconPapers.repec.org/RePEc:arx:papers:1908.02545.

Linton, O., Whang, Y.J., 2007. The quantilogram: With an application to evaluating directional predictability. Journal of Econometrics 141, 250 – 282. URL: http://www.sciencedirect.com/science/article/pii/S0304407607000152, doi:10.1016/j.jeconom.2007.01.004. semiparametric methods in econometrics.

Nelsen, R.B., 2006. An Introduction to Copulas -. 2nd ed., Springer, Berlin, Heidelberg.

Otneim, H., Tjøstheim, D., 2017. The locally Gaussian density estimator for multivariate data. Statistics and Computing 27, 1595–1616. URL: https://doi.org/10.1007/s11222-016-9706-6, doi:10.1007/s11222-016-9706-6.

Otneim, H., Tjøstheim, D., 2018. Conditional density estimation using the local Gaussian correlation. Statistics and Computing 28, 303–321. URL: http://dx.doi.org/10.1007/s11222-017-9732-z, doi:10.1007/s11222-017-9732-z.

Silvapulle, P., Granger, C., 2001. Large returns, conditional correlation and portfolio diversification: a value-at-risk approach. Quantitative Finance 1, 542–551. URL: https://doi.org/10.1080/713665877, doi:10.1080/713665877, arXiv:https://doi.org/10.1080/713665877.

Sklar, A., 1959. Fonctions de Répartition à $n$ dimensions et leurs Marges. Publications de l'Institut de Statistique de l'Université de Paris 8, 229–231.

Støve, B., Tjøstheim, D., 2014. Asymmetric dependence patterns in financial returns: An empirical investigation using local Gaussian correlation, in: Haldrup, N., Meitz, M., Saikkonen, P. (Eds.), Essays in Nonlinear Time Series Econometrics. Oxford University Press, Oxford. number 9780199679959 in OUP Catalogue, pp. 307–329.

Støve, B., Tjøstheim, D., Hufthammer, K.O., 2014. Using local gaussian correlation in a nonlinear re-examination of financial contagion. Journal of Empirical Finance 25, 62–82.

Székely, G.J., Rizzo, M.L., 2009. Brownian distance covariance. Ann. Appl. Stat. 3, 1236–1265. URL: https://doi.org/10.1214/09-AOAS312, doi:10.1214/09-AOAS312.

Teräsvirta, T., Tjøstheim, D., Granger, C.W.J., 2010. Modelling nonlinear economic time series. Oxford University Press, Oxford.

Tjøstheim, D., Hufthammer, K.O., 2013. Local Gaussian correlation: A new measure of dependence. Journal of Econometrics 172, 33 – 48. URL: http://www.sciencedirect.com/science/article/pii/S0304407612001741, doi:10.1016/j.jeconom.2012.08.001.

Tong, H., 1990. Non-linear Time Series: A Dynamical System Approach. Oxford University Press, Oxford.

Tukey, J.W., 1959. An introduction to the measurement of spectra, in: Grenander, U. (Ed.), Probability and Statistics, The Harald Cramér Volume, Almqvist and Wiksell, Stockholm, Sweden. pp. 300–330.

Wang, X., Hong, Y., 2018. Characteristic Function Based Testing for Conditional Independence: A Nonparametric Regression approach. Econometric Theory 34, 815–849. URL: https://doi.org/10.1017/S026646661700010X, doi:10.1017/S026646661700010X.

# SUPPLEMENTARY MATERIAL

This part contains the supplementary material to the paper *Nonlinear spectral analysis: A local Gaussian approach*. The asymptotic results for $\widehat{f}_{\boldsymbol{v}}^{m}(\omega)$ are presented in appendix A, appendix B contains the underlying asymptotic results for the parameters $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\bar{m}|\boldsymbol{b}}$, and a collection of technical details is given in appendix C.

A sensitivity analysis of the tuning parameters is given in appendix D, and some comments related to the selection of the tuning parameters are given in appendix E. Appendix F discusses issues related to sampling and resampling, including a sensitivity analysis of the block length $L$ for the slightly adjusted block bootstrap that is used in this paper.

Finally, appendix G contains some additional information about the examples used in the main document, and it does also include comments related to the reproducibility scripts that are contained in the `R`-package `localgaussSpec`.

## Appendix A: Asymptotic results for $\widehat{f}_{\boldsymbol{v}}^{m}(\omega)$

This appendix presents the asymptotic properties of $\widehat{f}_{\boldsymbol{v}}^{m}(\omega)$, the *m-truncated estimate of the local Gaussian spectral density*, i.e. the proof of theorem 2.8 is given here together with a theorem that covers the case when $\widehat{f}_{\boldsymbol{v}}^{m}(\omega)$ is complex-valued. The technical details needed for the proofs are covered in appendices B and C. Note that the theory is given for the general situation, i.e. it is not required that the time series under investigation should have been replaced with a pseudo-normalised version.

### A.1 The proof of theorem 2.8

*Proof.* The property that $f_{\boldsymbol{v}}(\omega)$ is a real-valued function when $\boldsymbol{v}$ lies on the diagonal was proved in lemma 2.3(d). The expression for $\widehat{f}_{\boldsymbol{v}}^{m}(\omega)$ from lemma 2.6 can by vectors be written as

$$\widehat{f}_{\boldsymbol{v}}^{m}(\omega) = 1 + 2 \cdot \boldsymbol{\Lambda}'_{m}(\omega) \cdot \widehat{\boldsymbol{P}}_{\boldsymbol{v}|m|\boldsymbol{b}}, \tag{A.1}$$

i.e. the sum can be expressed as the inner product of the two vectors

$$\boldsymbol{\Lambda}'_{m}(\omega) \coloneqq \left[\lambda_{m}(1) \cdot \cos\left(2\pi\omega \cdot 1\right), \ldots, \lambda_{m}(m) \cdot \cos\left(2\pi\omega \cdot m\right)\right], \tag{A.2a}$$

$$\widehat{\boldsymbol{P}}_{\boldsymbol{v}|m|\boldsymbol{b}} \coloneqq \left[\widehat{\rho}_{\boldsymbol{v}}(1|\boldsymbol{b}_{1}), \ldots, \widehat{\rho}_{\boldsymbol{v}}(m|\boldsymbol{b}_{m})\right]'. \tag{A.2b}$$

Since $\widehat{\rho}_{\boldsymbol{v}}(h|\boldsymbol{b}_{h})$ is one of the 5 estimated parameters $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}}(h|\boldsymbol{b}_{h})$ from the local Gaussian approximation (of the lag $h$ pairs) at the point $\boldsymbol{v}$,[1] it is clear that it is possible to write

[1] The properties of $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}}(h|\boldsymbol{b}_{h})$ was investigated in Tjøstheim and Hufthammer [2013]. A brief summary, with notation adjusted to fit the multivariate framework of the present paper, is given appendix B.1.2.

$\widehat{\rho}_{\boldsymbol{v}}(h|\boldsymbol{b}_h) = \boldsymbol{e}_5' \cdot \widehat{\boldsymbol{\theta}}_{\boldsymbol{v}}(h|\boldsymbol{b}_h)$, where $\boldsymbol{e}_5'$ is the unit vector that picks out $\widehat{\rho}_{\boldsymbol{v}}(h|\boldsymbol{b}_h)$ from $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}}(h|\boldsymbol{b}_h)$. The vectors $\left\{\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}}(h|\boldsymbol{b}_h)\right\}_{h=1}^{m}$ can be stacked on top of each other to give a joint parameter vector $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}}$, and it follows that the vector $\widehat{\boldsymbol{P}}_{\boldsymbol{v}|m|\boldsymbol{b}}$ can be expressed as $\widehat{\boldsymbol{P}}_{\boldsymbol{v}|m|\boldsymbol{b}} = \boldsymbol{E}_m' \cdot \widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}}$, where $\boldsymbol{E}_m'$ is the matrix that picks out the relevant components from $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}}$. It follows from this, and Brockwell and Davis [1986, Proposition 6.4.2, p. 211], that an asymptotic normality result for $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}}$ will give an asymptotic normality result for $\widehat{f}_{\boldsymbol{v}}^{m}(\omega)$. In particular, if a suitable scaling factor[2] $c_{n|m|\boldsymbol{b}}$ gives a $5m$-variate asymptotic normality result for $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}}$,

$$c_{n|m|\boldsymbol{b}} \cdot \left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}} - \boldsymbol{\theta}_{\boldsymbol{v}|\overline{m}}\right) \xrightarrow{d} \mathrm{N}\left(\boldsymbol{0}, \Sigma_{\boldsymbol{v}|\overline{m}}\right), \tag{A.3}$$

then a scaling factor $c'_{n|m|\boldsymbol{b}}$ can be found that gives a univariate asymptotic normality result for $\widehat{f}_{\boldsymbol{v}}^{m}(\omega)$,

$$c'_{n|m|\boldsymbol{b}} \cdot \left(\widehat{f}_{\boldsymbol{v}}^{m}(\omega) - f_{\boldsymbol{v}}(\omega)\right) \xrightarrow{d} \mathrm{N}(\boldsymbol{0}, \sigma_{\boldsymbol{v}}^2(\omega)), \tag{A.4}$$

where the variance $\sigma_{\boldsymbol{v}}^2(\omega)$ is a suitably scaled version of the limit of

$$\begin{aligned}\mathrm{Var}\left(\widehat{f}_{\boldsymbol{v}}^{m}(\omega)\right) &= 4 \cdot \mathrm{Var}\left(\boldsymbol{\Lambda}_m'(\omega) \cdot \boldsymbol{E}_m' \cdot \widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}}\right) \\ &= 4 \cdot \boldsymbol{\Lambda}_m'(\omega) \cdot \boldsymbol{E}_m' \cdot \mathrm{Var}\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}}\right) \cdot \boldsymbol{E}_m \cdot \boldsymbol{\Lambda}_m(\omega).\end{aligned} \tag{A.5}$$

The asymptotic normality required in eq. (A.3) follows from theorem B.22 (page 33), i.e. the scaling factor $c_{n|m|\boldsymbol{b}}$ will be $\sqrt{n(b_1 b_2)^3}$, whereas the asymptotic covariance matrix $\Sigma_{\boldsymbol{v}|\overline{m}}$ can be written as the direct sum of the covariance matrices for $\sqrt{n(b_1 b_2)^3} \cdot \widehat{\boldsymbol{\theta}}_{\boldsymbol{v}}(h|\boldsymbol{b}_h)$, i.e.

$$\mathrm{Var}\left(\sqrt{n(b_1 b_2)^3} \cdot \widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}}\right) = \bigoplus_{h=1}^{m} \mathrm{Var}\left(\sqrt{n(b_1 b_2)^3} \cdot \widehat{\boldsymbol{\theta}}_{\boldsymbol{v}}(h|\boldsymbol{b}_h)\right), \tag{A.6}$$

from which a simple calculation gives

$$\mathrm{Var}\left(\sqrt{n(b_1 b_2)^3} \cdot \widehat{f}_{\boldsymbol{v}}^{m}(\omega)\right) = 4 \cdot \sum_{h=1}^{m} \lambda_m^2(h) \cdot \cos^2(2\pi\omega h) \cdot \mathrm{Var}\left(\sqrt{n(b_1 b_2)^3} \cdot \widehat{\rho}_{\boldsymbol{v}}(h|\boldsymbol{b}_h)\right). \tag{A.7}$$

[2] $c_{n|m|\boldsymbol{b}}$ must be a function of $n$, $m$ and $\{\boldsymbol{b}_h\}_{h=1}^{m}$, such that $c_{n|m|\boldsymbol{b}} \to \infty$ when $n \to \infty$, $m \to \infty$ and $\boldsymbol{b}_h \to \boldsymbol{0}^+$.

From this it is clear that the scaling factor $c_{n|m|\boldsymbol{b}}$ requires an additional scaling with $\sqrt{1/m}$ in order to include the averaging factor $1/m$ for the sum in eq. (A.7). Thus, $c'_{n|m|\boldsymbol{b}} = \sqrt{n(b_1 b_2)^3/m}$, which completes the proof. □

Some care must be taken formally with regard to the limiting $5m$-variate normal distribution in eq. (A.3), since it has to be interpreted as something that is approximately valid for large (but finite) values of the truncation point $m$. The univariate normal distribution in eq. (A.4) is the one of interest, and this will under the required assumptions be well defined in the limit.

## A.2 The complex-valued case

**Theorem A.1** (Complex-valued case)**.** *If the local Gaussian spectral density $f_{\boldsymbol{v}}(\omega)$ is a complex valued function for a point $\boldsymbol{v} = (v_1, v_2)$, i.e. $f_{\boldsymbol{v}}(\omega) = c_{\boldsymbol{v}}(\omega) - iq_{\boldsymbol{v}}(\omega)$, with $q_{\boldsymbol{v}}(\omega) \not\equiv 0$, then, under assumptions 2.1 to 2.3, the components $\widehat{c}_{\boldsymbol{v}}{}^m(\omega)$ and $\widehat{q}_{\boldsymbol{v}}{}^m(\omega)$ of the $m$-truncated estimate $\widehat{f}_{\boldsymbol{v}}^m(\omega)$ will, when $\omega \notin \frac{1}{2}\cdot\mathbb{Z} := \{\ldots, -1, -\frac{1}{2}, 0, \frac{1}{2}, 1, \ldots\}$, be jointly asymptotically normally distributed as given below.*

$$\sqrt{n(b_1 b_2)^3/m} \cdot \left( \begin{bmatrix} \widehat{c}_{\boldsymbol{v}}{}^m(\omega) \\ \widehat{q}_{\boldsymbol{v}}{}^m(\omega) \end{bmatrix} - \begin{bmatrix} c_{\boldsymbol{v}}(\omega) \\ q_{\boldsymbol{v}}(\omega) \end{bmatrix} \right) \xrightarrow{d} \mathrm{N}\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{c:\boldsymbol{v}}(\omega) & 0 \\ 0 & \sigma^2_{q:\boldsymbol{v}}(\omega) \end{bmatrix} \right), \tag{A.8}$$

*where the variances $\sigma^2_{c:\boldsymbol{v}}(\omega)$ and $\sigma^2_{q:\boldsymbol{v}}(\omega)$ are given by*

$$\sigma^2_{c:\boldsymbol{v}}(\omega) = \lim_{m\to\infty} \frac{1}{m} \sum_{h=1}^{m} \lambda_m^2(h) \cdot \cos^2(2\pi\omega h) \cdot \{\tilde{\sigma}^2_{\boldsymbol{v}}(h) + \tilde{\sigma}^2_{\check{\boldsymbol{v}}}(h)\} \tag{A.9a}$$

$$\sigma^2_{q:\boldsymbol{v}}(\omega) = \lim_{m\to\infty} \frac{1}{m} \sum_{h=1}^{m} \lambda_m^2(h) \cdot \sin^2(2\pi\omega h) \cdot \{\tilde{\sigma}^2_{\boldsymbol{v}}(h) + \tilde{\sigma}^2_{\check{\boldsymbol{v}}}(h)\}, \tag{A.9b}$$

*with $\tilde{\sigma}^2_{\boldsymbol{v}}(h)$ and $\tilde{\sigma}^2_{\check{\boldsymbol{v}}}(h)$ related to respectively $\widehat{\rho}_{\boldsymbol{v}}(h|\boldsymbol{b}_h)$ and $\widehat{\rho}_{\check{\boldsymbol{v}}}(h|\boldsymbol{b}_h)$ as given in theorem 2.8.*

*The component $\widehat{q}_{\boldsymbol{v}}{}^m(\omega)$ is identical to 0 when $\omega \in \frac{1}{2}\cdot\mathbb{Z}$, and for these frequencies the following asymptotic result holds under the given assumptions*

$$\sqrt{n(b_1 b_2)^3/m} \cdot \left( \widehat{f}_{\boldsymbol{v}}^m(\omega) - f_{\boldsymbol{v}}(\omega) \right) \xrightarrow{d} \mathrm{N}(0, \sigma^2_{c:\boldsymbol{v}}(\omega)). \tag{A.10}$$

*Proof.* The case $\omega \in \frac{1}{2}\cdot\mathbb{Z}$ can be proved by the exact same argument that was used in the proof of theorem 2.8, whereas the general case requires a bivariate extension of that proof.

In particular, when the proof of theorem 2.8 is used on $\widehat{c}_{\boldsymbol{v}}^{\,m}(\omega)$ and $\widehat{q}_{\boldsymbol{v}}^{\,m}(\omega)$, it follows that they can be written as

$$\widehat{c}_{\boldsymbol{v}}^{\,m}(\omega) = 1 + \boldsymbol{\Lambda}'_{c|m}(\omega) \cdot \widehat{\boldsymbol{P}}_{\boldsymbol{v}|m|\boldsymbol{b}} + \boldsymbol{\Lambda}'_{c|m}(\omega) \cdot \widehat{\boldsymbol{P}}_{\breve{\boldsymbol{v}}|m|\boldsymbol{b}} = 1 + \boldsymbol{\Lambda}'_{c|\underline{\overline{m}}}(\omega) \cdot \widehat{\boldsymbol{P}}_{\boldsymbol{v}|\underline{\overline{m}}|\boldsymbol{b}} \tag{A.11a}$$

$$\widehat{q}_{\boldsymbol{v}}^{\,m}(\omega) = 0 + \boldsymbol{\Lambda}'_{q|m}(\omega) \cdot \widehat{\boldsymbol{P}}_{\boldsymbol{v}|m|\boldsymbol{b}} - \boldsymbol{\Lambda}'_{q|m}(\omega) \cdot \widehat{\boldsymbol{P}}_{\breve{\boldsymbol{v}}|m|\boldsymbol{b}} = 0 + \boldsymbol{\Lambda}'_{q|\underline{\overline{m}}}(\omega) \cdot \widehat{\boldsymbol{P}}_{\boldsymbol{v}|\underline{\overline{m}}|\boldsymbol{b}}, \tag{A.11b}$$

where $\boldsymbol{\Lambda}'_{c|m}(\omega)$ and $\boldsymbol{\Lambda}'_{q|m}(\omega)$ are the coefficient vectors containing respectively the cosines and sines, where $\widehat{\boldsymbol{P}}_{\boldsymbol{v}|m|\boldsymbol{b}}$ and $\widehat{\boldsymbol{P}}_{\breve{\boldsymbol{v}}|m|\boldsymbol{b}}$ contains the estimated correlations corresponding to $\boldsymbol{v}$ and $\breve{\boldsymbol{v}}$ for the lags under consideration, and where the length $2m$ vectors $\boldsymbol{\Lambda}'_{c|\underline{\overline{m}}}(\omega)$, $\boldsymbol{\Lambda}'_{q|\underline{\overline{m}}}(\omega)$ and $\widehat{\boldsymbol{P}}_{\boldsymbol{v}|\underline{\overline{m}}|\boldsymbol{b}}$ are defined in the obvious manner in order to get a more compact notation. Following the same line of argument as in the proof of theorem 2.8, it follows that $\widehat{\boldsymbol{P}}_{\boldsymbol{v}|\underline{\overline{m}}|\boldsymbol{b}} = (\boldsymbol{E}'_m \oplus \boldsymbol{E}'_m) \cdot \widehat{\boldsymbol{\Theta}}_{\underline{\overline{m}}|\boldsymbol{b}}(\boldsymbol{v}, \breve{\boldsymbol{v}})$, where $\widehat{\boldsymbol{\Theta}}_{\underline{\overline{m}}|\boldsymbol{b}}(\boldsymbol{v}, \breve{\boldsymbol{v}})$ is the full set of estimated parameters from the local Gaussian approximations at $\boldsymbol{v}$ and $\breve{\boldsymbol{v}}$ for the lags under consideration,[3] and where $(\boldsymbol{E}'_m \oplus \boldsymbol{E}'_m)$ is the matrix that picks out the relevant autocorrelations.

Based upon this, it follows that the target of interest can be written as

$$\begin{bmatrix} \widehat{c}_{\boldsymbol{v}}^{\,m}(\omega) \\ \widehat{q}_{\boldsymbol{v}}^{\,m}(\omega) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Lambda}'_{c|\underline{\overline{m}}}(\omega) \\ \boldsymbol{\Lambda}'_{q|\underline{\overline{m}}}(\omega) \end{bmatrix} \cdot (\boldsymbol{E}'_m \oplus \boldsymbol{E}'_m) \cdot \widehat{\boldsymbol{\Theta}}_{\underline{\overline{m}}|\boldsymbol{b}}(\boldsymbol{v}, \breve{\boldsymbol{v}}) \,, \tag{A.12}$$

which together with the asymptotic normality result from theorem B.23, i.e.

$$\sqrt{n(b_1 b_2)^3} \cdot \Big( \widehat{\boldsymbol{\Theta}}_{\underline{\overline{m}}|\boldsymbol{b}}(\boldsymbol{v}, \breve{\boldsymbol{v}}) - \boldsymbol{\Theta}_{\underline{\overline{m}}}(\boldsymbol{v}, \breve{\boldsymbol{v}}) \Big) \xrightarrow{d} \mathrm{N}\big(\boldsymbol{0}, \Sigma_{\boldsymbol{v}|\underline{\overline{m}}} \oplus \Sigma_{\breve{\boldsymbol{v}}|\underline{\overline{m}}}\big) \,, \tag{A.13}$$

gives the result when the arguments in the proof of theorem 2.8 are applied to the present setup. Note that the requirement $\omega \notin \frac{1}{2} \cdot \mathbb{Z}$ is needed in order to ensure that the variance $\sigma^2_{q:\boldsymbol{v}}(\omega)$ is different from 0, which is needed in order for Brockwell and Davis [1986, Proposition 6.4.2, p. 211] to be valid in this case. □

## A.3 The finite sample case and the variance of $\widehat{f}_{\boldsymbol{v}}^{\,m}(\omega)$

The variance of the estimated local Gaussian spectral density $\widehat{f}_{\boldsymbol{v}}^{\,m}(\omega)$, as seen in eq. (A.5), is a function of both the point $\boldsymbol{v}$ and the frequency $\omega$. It is with regard to this of interest to note that the variance $\sigma^2_{\boldsymbol{v}}(\omega)$ is symmetric around $\omega = \frac{1}{4}$, and it attains its highest values when $\omega \in \{0, \frac{1}{2}\}$. This symmetry is a consequence of the fact that all the correlation terms are asymptotically negligible.

[3] The vector $\widehat{\boldsymbol{\Theta}}_{\underline{\overline{m}}|\boldsymbol{b}}(\boldsymbol{v}, \breve{\boldsymbol{v}})$ can be expressed as a combination of $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}}$ and $\widehat{\boldsymbol{\theta}}_{\breve{\boldsymbol{v}}|\overline{m}|\boldsymbol{b}}$, where $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}}$ is the parameter vector from the proof of theorem 2.8.

The correlation-terms are however still present in the $m$-truncated case, and this changes the situation a bit. To clarify: The correlation terms will depend on the frequency $\omega$ trough the functions $\cos(2\pi\omega k)\cdot\cos(2\pi\omega\ell)$, and these functions are in general not symmetrical around $\omega=\frac{1}{4}$. For $\omega=0$ all these products are equal to 1, whereas the value for $\omega=\frac{1}{2}$ will be given by $\cos(2\pi\omega k)\cdot\cos(2\pi\omega\ell)=(-1)^{k+\ell}$. The consequence of this is that the highest value of this variance is obtained at $\omega=0$ — which in particular was evident in the plots related to the apARCH-model and the `dmpb`-data, cf. figs. 9 and 11 in the main document, where a 'trumpet shape' could be seen for the pointwise confidence intervals near $\omega=0$.

# Appendix B: Asymptotic results for $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}}$

This section will investigate the asymptotic properties of the parameter vector $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}}$, that is used in the proof of theorem 2.8. The proof is similar in spirit to the one used in Tjøstheim and Hufthammer [2013] for the asymptotic investigation of the parameter vectors $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}}(h|\boldsymbol{b}_h)$, i.e. the Klimko-Nelson penalty function approach will be used to derive the desired result.

Appendix B.1 explains the Klimko-Nelson approach and shows how a local penalty function for the present case can be constructed based on the local penalty function encountered in Tjøstheim and Hufthammer [2013]. Appendix B.2 verifies the fourth of the requirements needed for the Klimko-Nelson approach, and the asymptotic results for $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}}$ are collected in appendix B.3.

The asymptotic investigation requires several indices in order to keep track of the different components, and to simplify references to $\boldsymbol{v}$ and $\boldsymbol{b}$ will whenever possible be suppressed from the notation.

## B.1 Local penalty functions and the Klimko-Nelson approach

Tjøstheim and Hufthammer [2013] used a local penalty function to define the *local Gaussian correlation* $\rho_{\boldsymbol{v}}$ as a new *local measure of dependence* at a point $\boldsymbol{v}$, and then used the approach formalised in Klimko and Nelson [1978], to investigate the asymptotic properties of $\widehat{\rho}_{\boldsymbol{v}}$. The *local Gaussian spectral density* $f_{\boldsymbol{v}}(\omega)$ is based on the local Gaussian autocorrelations $\rho_{\boldsymbol{v}}(h)$, and the asymptotic properties of the estimates $\widehat{f}^{m}_{\boldsymbol{v}}(\omega)$ are thus closely connected to the asymptotic properties of $\widehat{\rho}_{\boldsymbol{v}}(h)$.

The Klimko-Nelson approach shows how the asymptotic properties of *an estimate of the parameters of a penalty function* $Q$ can be expressed relative to the asymptotic properties of (entities related to) the penalty function itself. This result plays a pivotal role in the present analysis, and it has thus been included in appendix B.1.1.

Appendix B.1.2 presents the bivariate definitions and results from Tjøstheim and Hufthammer [2013], with the notational modifications that are needed in order to make it fit into

the multivariate approach in the present paper. The bivariate penalty functions $Q_{h:n}$ from Tjøstheim and Hufthammer [2013] will be used as building blocks for the new penalty function.

### B.1.1 The Klimko-Nelson approach

The following presentation is based on Taniguchi and Kakizawa [2000, Th. 3.2.23].

Let $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ be an $m$-variate strictly stationary and ergodic process that satisfies the requirement $\mathrm{E}\left[\|\boldsymbol{X}_t\|^2\right] < \infty$. Consider a general real valued penalty function $Q_n = Q_n(\boldsymbol{\theta}) = Q_n(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \boldsymbol{\theta})$, which should depend upon $n$ observations $\{\boldsymbol{X}_t\}_{i=1}^n$ and a parameter vector $\boldsymbol{\theta}$ that lies in an open set $\boldsymbol{\Theta} \in \mathbb{R}^p$, and let the true value of the parameter be denoted by $\boldsymbol{\theta}^\circ$. Add the requirement that $Q_n$ must be twice continuously differentiable with respect to $\boldsymbol{\theta}$ a.e. in a neighbourhood $\mathcal{N}$ of $\boldsymbol{\theta}^\circ$, such that the following Taylor expansion is valid (in the neighbourhood $\mathcal{N}$) for $\|\boldsymbol{\theta} - \boldsymbol{\theta}^\circ\| < \delta$,

$$
\begin{aligned}
Q_n(\boldsymbol{\theta}) &= Q_n(\boldsymbol{\theta}^\circ) + (\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)' \frac{\partial}{\partial \boldsymbol{\theta}} Q_n(\boldsymbol{\theta}^\circ) + \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)' \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} Q_n(\boldsymbol{\theta}^\circ) (\boldsymbol{\theta} - \boldsymbol{\theta}^\circ) \\
&\quad + \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)' \left\{ \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} Q_n(\boldsymbol{\theta}^*) - \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} Q_n(\boldsymbol{\theta}^\circ) \right\} (\boldsymbol{\theta} - \boldsymbol{\theta}^\circ) \qquad \text{(B.1a)} \\
&= Q_n(\boldsymbol{\theta}^\circ) + (\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)' \frac{\partial}{\partial \boldsymbol{\theta}} Q_n(\boldsymbol{\theta}^\circ) + \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)' V_n (\boldsymbol{\theta} - \boldsymbol{\theta}^\circ) \\
&\quad + \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)' T_n(\boldsymbol{\theta}^*) (\boldsymbol{\theta} - \boldsymbol{\theta}^\circ) \qquad \text{(B.1b)}
\end{aligned}
$$

where $V_n$ and $T_n(\boldsymbol{\theta}^*)$ are defined in the obvious manner, with $\boldsymbol{\theta}^* = \boldsymbol{\theta}^*(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \boldsymbol{\theta})$ an intermediate point between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^\circ$ (determined by the mean value theorem).

**Theorem B.1** (Klimko-Nelson, Klimko and Nelson [1978])**.** *Assume that* $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ *and* $Q_n$ *are such that as* $n \to \infty$

(A1) $n^{-1}(\partial/\partial\boldsymbol{\theta})Q_n(\boldsymbol{\theta}^\circ) \xrightarrow{a.s.} \mathbf{0}$,

(A2) $n^{-1}V_n \xrightarrow{a.s.} V$, *where* $V$ *is a* $p \times p$ *positive definite matrix, and*

(A3) *for* $j, k = 1, \ldots, p$

$$
\lim_{n\to\infty} \sup_{\delta\to 0} (n\delta)^{-1} \left| T_n\{\boldsymbol{\theta}^*\}_{jk} \right| < \infty \qquad a.s. \qquad \text{(B.2)}
$$

*where* $T_n\{\boldsymbol{\theta}^*\}_{jk}$ *is the* $(j,k)$*th component of* $T_n\{\boldsymbol{\theta}^*\}$.

*Then there exists a sequence of estimators* $\widehat{\boldsymbol{\theta}}_n = \left(\widehat{\theta}_1, \ldots, \widehat{\theta}_p\right)'$, *such that* $\widehat{\boldsymbol{\theta}}_n \xrightarrow{a.s.} \boldsymbol{\theta}^\circ$, *and for any* $\epsilon > 0$, *there exists an event* $E$ *with* $P(E) > 1 - \epsilon$ *and an* $n^\circ$ *such that on* $E$, *for* $n > n^\circ$, $(\partial/\partial\boldsymbol{\theta})Q_n(\widehat{\boldsymbol{\theta}}_n) = \mathbf{0}$ *and* $Q_n$ *attains a relative minimum at* $\widehat{\boldsymbol{\theta}}_n$. *Furthermore, if*

(A4) $n^{-1/2}(\partial/\partial\boldsymbol{\theta})Q_{n}(\boldsymbol{\theta}^{\circ}) \xrightarrow{d} \mathrm{N}(\mathbf{0}, W)$
*then*

$$n^{1/2}(\widehat{\boldsymbol{\theta}}_{n} - \boldsymbol{\theta}^{\circ}) \xrightarrow{d} \mathrm{N}(\mathbf{0}, V^{-1}WV^{-1})\,. \tag{B.3}$$

### B.1.2 The bivariate penalty functions

This section will translate the bivariate results from Tjøstheim and Hufthammer [2013] into the present multivariate framework, and these bivariate components will then be used to define a new penalty function in appendix B.1.3.

The main idea from Tjøstheim and Hufthammer [2013] is to use bivariate Gaussian densities $\psi\big(\boldsymbol{y}_{h};\boldsymbol{\theta}_{\boldsymbol{v}|h}\big)$ to approximate the bivariate densities $g_{h}(\boldsymbol{y}_{h})$ at a point $\boldsymbol{v}$, where $\boldsymbol{\theta}_{\boldsymbol{v}|h} = \big[\theta_{\boldsymbol{v}|h:1}, \ldots, \theta_{\boldsymbol{v}|h:5}\big]'$ is the five dimensional parameter-vector of the bivariate Gaussian distribution. The point $\boldsymbol{v}$ will be fixed for the remainder of this discussion, and it will henceforth be dropped from the notation for the parameters, i.e. $\boldsymbol{\theta}_{h}$ should always be understood as $\boldsymbol{\theta}_{\boldsymbol{v}|h}$.

The local investigation requires a bandwidth vector $\boldsymbol{b} = (b_{1}, b_{2})$ and a kernel function $K(\boldsymbol{w})$, which is used to define $K_{h:\boldsymbol{b}}(\boldsymbol{y}_{h} - \boldsymbol{v}) \coloneqq \frac{1}{b_{1}b_{2}} K\Big(\frac{y_{h}-v_{1}}{b_{1}}, \frac{y_{0}-v_{2}}{b_{2}}\Big)$, which in turn is used in the following local approximation around $\boldsymbol{v}$,

$$q_{h:\boldsymbol{b}} \coloneqq \int_{\mathbb{R}^{2}} K_{h:\boldsymbol{b}}(\boldsymbol{y}_{h} - \boldsymbol{v}) \left[\psi(\boldsymbol{y}_{h};\boldsymbol{\theta}_{h}) - g_{h}(\boldsymbol{y}_{h}) \log \psi(\boldsymbol{y}_{h};\boldsymbol{\theta}_{h})\right] \mathrm{d}\boldsymbol{y}_{h}, \tag{B.4}$$

a minimiser of which should satisfy the vector equation

$$\int_{\mathbb{R}^{2}} K_{h:\boldsymbol{b}}(\boldsymbol{y}_{h} - \boldsymbol{v})\, \boldsymbol{u}_{h}(\boldsymbol{y}_{h};\boldsymbol{\theta}_{h}) \left[\psi(\boldsymbol{y}_{h};\boldsymbol{\theta}_{h}) - g_{h}(\boldsymbol{y}_{h})\right] \mathrm{d}\boldsymbol{y}_{h} = \mathbf{0}, \tag{B.5}$$

where $\boldsymbol{u}_{h}(\boldsymbol{y}_{h};\boldsymbol{\theta}_{h}) \coloneqq \boldsymbol{\nabla}_{h} \log \psi(\boldsymbol{y}_{h};\boldsymbol{\theta}_{h})$ is the score function of $\psi(\boldsymbol{y}_{h};\boldsymbol{\theta}_{h})$ (with $\boldsymbol{\nabla}_{h} \coloneqq \partial/\partial\boldsymbol{\theta}_{h}$). Under the assumption that there is a bandwidth $\boldsymbol{b}_{0}$ such that there exists a minimiser $\boldsymbol{\theta}_{h:\boldsymbol{b}}$ of eq. (B.4) which satisfies eq. (B.5) for any $\boldsymbol{b}$ with $\mathbf{0} < \boldsymbol{b} < \boldsymbol{b}_{0}$,[4] this $\boldsymbol{\theta}_{h:\boldsymbol{b}}$ will be referred to as the population value for the given bandwidth $\boldsymbol{b}$.

Equation (B.4) is a special case of a tool that Hjort and Jones [1996] introduced in order to perform *locally parametric nonparametric density estimation*, but (as was done in Tjøstheim and Hufthammer [2013]) it can also be used to define and estimate local Gaussian parameters — whose asymptotic properties can be investigated by means of a local penalty function $Q_{h:n}(\boldsymbol{\theta}_{h})$, to be described below, and the Klimko-Nelson approach.

[4] Inequalities involving vectors are to be interpreted in a component-wise manner.

For a sample of size $n$ from $\{\boldsymbol{Y}_{h:t}\}_{t\in\mathbb{Z}}$, the following $M$-estimator[5] will be used, which (due to the ergodicity implied by assumption 2.1(a)) will converge towards the penalty function $q_{h:\boldsymbol{b}}$,

$$\begin{aligned} L_{h:n}(\boldsymbol{\theta}_h) &\coloneqq L_{h:n}(\boldsymbol{Y}_{h:1},\ldots,\boldsymbol{Y}_{h:n};\boldsymbol{\theta}_h) \\ &\coloneqq n^{-1}\sum_{t=1}^{n} K_{h:\boldsymbol{b}}(\boldsymbol{Y}_{h:t}-\boldsymbol{v})\log\psi(\boldsymbol{Y}_{h:t};\boldsymbol{\theta}_h) - \int_{\mathbb{R}^2} K_{h:\boldsymbol{b}}(\boldsymbol{y}_h-\boldsymbol{v})\,\psi(\boldsymbol{y}_h;\boldsymbol{\theta}_h)\,\mathrm{d}\boldsymbol{y}_h. \end{aligned} \tag{B.6}$$

The local penalty function from Tjøstheim and Hufthammer [2013] can be described as

$$\begin{aligned} Q_{h:n}(\boldsymbol{\theta}_h) &\coloneqq Q_{h:n}(\boldsymbol{Y}_{h:1},\ldots,\boldsymbol{Y}_{h:n};\boldsymbol{\theta}_h) \coloneqq -nL_{h:n}(\boldsymbol{\theta}_h) \\ &= -\sum_{t=1}^{n} K_{h:\boldsymbol{b}}(\boldsymbol{Y}_{h:t}-\boldsymbol{v})\log\psi(\boldsymbol{Y}_{h:t};\boldsymbol{\theta}_h) + n\int_{\mathbb{R}^2} K_{h:\boldsymbol{b}}(\boldsymbol{y}_h-\boldsymbol{v})\,\psi(\boldsymbol{y}_h;\boldsymbol{\theta}_h)\,\mathrm{d}\boldsymbol{y}_h, \end{aligned} \tag{B.7}$$

and it remains to write out how the different components in appendix B.1.1 looks like for this particular penalty function. A central component is the vector of partial derivatives, which by the score function $\boldsymbol{u}_h(\boldsymbol{y}_h;\boldsymbol{\theta}_h)$ can be given as,

$$\boldsymbol{\nabla}_h Q_{h:n}(\boldsymbol{\theta}_h) = -\sum_{t=1}^{n}\left[K_{h:\boldsymbol{b}}(\boldsymbol{Y}_{h:t}-\boldsymbol{v})\,\boldsymbol{u}_h(\boldsymbol{Y}_{h:t};\boldsymbol{\theta}_h) - \int_{\mathbb{R}^2} K_{h:\boldsymbol{b}}(\boldsymbol{y}_h-\boldsymbol{v})\,\boldsymbol{u}_h(\boldsymbol{y}_h;\boldsymbol{\theta}_h)\,\psi(\boldsymbol{y}_h;\boldsymbol{\theta}_h)\,\mathrm{d}\boldsymbol{y}_h\right]. \tag{B.8}$$

Note that the expectation of the bracketed expression in the sum gives the left hand side of eq. (B.5), which implies that the expectation will be $\mathbf{0}$ when $\boldsymbol{\nabla}_h Q_{h:n}(\boldsymbol{\theta}_h)$ is evaluated at the population value $\boldsymbol{\theta}_{h:\boldsymbol{b}}$.

Given a bandwidth $\boldsymbol{b}$ which is small enough to ensure a unique solution $\boldsymbol{\theta}_{h:\boldsymbol{b}}$, the next part of interest is the Taylor expansion of order two in a neighbourhood $\mathcal{N}_h \coloneqq \{\boldsymbol{\theta}_h : |\boldsymbol{\theta}_h - \boldsymbol{\theta}_{h:\boldsymbol{b}}| < \delta\}$ of $\boldsymbol{\theta}_{h:\boldsymbol{b}}$, i.e.

$$\begin{aligned} Q_{h:n}(\boldsymbol{\theta}_h) = Q_{h:n}(\boldsymbol{\theta}_{h:\boldsymbol{b}}) &+ [\boldsymbol{\theta}_h-\boldsymbol{\theta}_{h:\boldsymbol{b}}]'\,\boldsymbol{\nabla}_h Q_{h:n}(\boldsymbol{\theta}_{h:\boldsymbol{b}}) + \frac{1}{2}[\boldsymbol{\theta}_h-\boldsymbol{\theta}_{h:\boldsymbol{b}}]'\,V_{h:\boldsymbol{b}:n}\,[\boldsymbol{\theta}_h-\boldsymbol{\theta}_{h:\boldsymbol{b}}] \\ &+ \frac{1}{2}[\boldsymbol{\theta}_h-\boldsymbol{\theta}_{h:\boldsymbol{b}}]'\,T_{h:\boldsymbol{b}:n}\,[\boldsymbol{\theta}_h-\boldsymbol{\theta}_{h:\boldsymbol{b}}], \end{aligned} \tag{B.9a}$$

[5] The entity $L_{h:n}(\boldsymbol{\theta}_h)$ can for independent observations be thought of as a *local log-likelihood* or a *local kernel-smoothed log-likelihood*, see Hjort and Jones [1996, Section 2-3] for details. In the realm of time series, where the observations are dependent, it is according to Tjøstheim and Hufthammer [2013, page 36] better to interpret it as an $M$-estimation penalty function

where

$$V_{h:\boldsymbol{b}:n} \coloneqq V_{h:\boldsymbol{b}:n}(\boldsymbol{\theta}_{h:\boldsymbol{b}}) \coloneqq \boldsymbol{\nabla}_h \boldsymbol{\nabla}_h' Q_{h:n}(\boldsymbol{\theta}_{h:\boldsymbol{b}}) \,, \tag{B.9b}$$

$$T_{h:\boldsymbol{b}:n} \coloneqq T_{h:\boldsymbol{b}:n}(\boldsymbol{\theta}_h^*, \boldsymbol{\theta}_{h:\boldsymbol{b}}) \coloneqq \boldsymbol{\nabla}_h \boldsymbol{\nabla}_h' Q_{h:n}(\boldsymbol{\theta}_h^*) - \boldsymbol{\nabla}_h \boldsymbol{\nabla}_h' Q_{h:n}(\boldsymbol{\theta}_{h:\boldsymbol{b}}) \,, \tag{B.9c}$$

with $\boldsymbol{\theta}_h^*$ an intermediate point between $\boldsymbol{\theta}_h$ and $\boldsymbol{\theta}_{h:\boldsymbol{b}}$, again determined by the mean value theorem.

With the preceding definitions, Tjøstheim and Hufthammer [2013, theorem 1] investigated the case where the bandwidth $\boldsymbol{b}$ was fixed as $n \to \infty$, i.e. items *(A1)* to *(A4)* of theorem B.1 was verified in order to obtain the following result for the estimated local Gaussian parameters $\widehat{\boldsymbol{\theta}}_{h:n}$; for every $\epsilon > 0$ there exists an event $A_h$ (possibly depending on the point $\boldsymbol{v}$) with $\mathrm{P}(A_h^c) < \epsilon$, such that there exists a sequence of estimators $\widehat{\boldsymbol{\theta}}_{h:n}$ that converges almost surely to $\boldsymbol{\theta}_{h:\boldsymbol{b}}$ (the minimiser of $q_{h:\boldsymbol{b}}$ from eq. (B.4)). And, moreover, the following asymptotic behaviour is observed

$$(n b_1 b_2)^{1/2} \left( \widehat{\boldsymbol{\theta}}_{h:n} - \boldsymbol{\theta}_{h:\boldsymbol{b}} \right) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \Sigma_{h:\boldsymbol{b}}) \,, \tag{B.10}$$

where $\Sigma_{h:\boldsymbol{b}} \coloneqq V_{h:\boldsymbol{b}}^{-1} W_{h:\boldsymbol{b}} V_{h:\boldsymbol{b}}^{-1}$ with $W_{h:\boldsymbol{b}}$ the matrix occurring in item *(A4)* of theorem B.1.

The situation when $\boldsymbol{b} \to \mathbf{0}^+$ as $n \to \infty$ requires some extra care since the presence of the kernel function $K_{h:\boldsymbol{b}}(\boldsymbol{w})$ in $Q_{h:n}(\boldsymbol{\theta}_h)$, see eq. (B.7), gives limiting matrices of $V_{h:\boldsymbol{b}}$ and $W_{h:\boldsymbol{b}}$ of rank one. The details are covered in theorems 2 and 3 in Tjøstheim and Hufthammer [2013, p. 39-40], which ends out with the following adjusted version of eq. (B.10), where $n$ and $\boldsymbol{b} = (b_1, b_2)$ are such that $\log n / n(b_1 b_2)^5 \to 0$,

$$\left( n \left( b_1 b_2 \right)^3 \right)^{1/2} \left( \widehat{\boldsymbol{\theta}}_{h:n} - \boldsymbol{\theta}_h^\circ \right) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \Sigma_h^\circ) \,, \tag{B.11}$$

where $\boldsymbol{\theta}_h^\circ$ is the $\boldsymbol{b} \to \mathbf{0}^+$ value of $\boldsymbol{\theta}_{h:\boldsymbol{b}}$ and where the limiting matrix $\Sigma_h^\circ$ is a $(b_1 b_2)^2$-rescaled version of matrices *related to* the matrices $V_{h:\boldsymbol{b}}$ and $W_{h:\boldsymbol{b}}$, see the discussion in Tjøstheim and Hufthammer [2013] for details.

### B.1.3 A new penalty function

The proof of theorem 2.8 requires an asymptotic result for the parameter vector $\widehat{\boldsymbol{\theta}}_{n|\overline{m}|\boldsymbol{b}}$, which was obtained by combining $m$ parameter vectors corresponding to the bivariate lag $h$ pairs $\left(Y_{t+h}, Y_t\right)$ for $h = 1, \ldots, m$. This section will show how a penalty function for $\widehat{\boldsymbol{\theta}}_{n|\overline{m}|\boldsymbol{b}}$ can be constructed based on the bivariate penalty functions $Q_{h:n}$ defined in appendix B.1.2. The indices $n$ and $\boldsymbol{b}$ will for notational simplicity be suppressed from the notation, and only $\boldsymbol{\theta}_{\overline{m}}$ will henceforth be used.

An analysis akin to the one in Theorem 1 of Tjøstheim and Hufthammer [2013] will be performed in this section, i.e. the asymptotic situation will be investigated for the simple case where the truncation $m$ and the bandwidth $\boldsymbol{b}$ both are fixed as $n \to \infty$. The proof that the new penalty function satisfies the four requirements items *(A1)* to *(A4)* of theorem B.1 can then be based upon corresponding components of the proof of Theorem 1 from Tjøstheim and Hufthammer [2013].

The general case, where $m \to \infty$ and $\boldsymbol{b} \to \mathbf{0}^+$ when $n \to \infty$, can recycle the arguments given here for the requirements in items *(A1)* to *(A3)*, but extra work is needed for the requirement given in item *(A4)*. The details needed for item *(A4)* will be covered in appendix B.2.

With regard to the construction of the new penalty function, the main observation of interest is that the $Q_{h:n}(\boldsymbol{\theta}_h)$ from appendix B.1.2 was defined for bivariate time series $\{\boldsymbol{Y}_{h:t}\}_{t\in\mathbb{Z}}$, whereas the new penalty function will be defined for the $(m+1)$-variate time series $\{\boldsymbol{Y}_{\overline{m}:t}\}_{t\in\mathbb{Z}}$. The first step is to extend the penalty functions $Q_{h:n}$, $h = 1, \ldots, m$ from expression based on $\boldsymbol{Y}_{h:t}$ to expressions based on $\boldsymbol{Y}_{\overline{m}:t}$, but this is trivial since the bivariate functions occurring in the definition of $Q_{h:n}(\boldsymbol{\theta}_h)$ can be extended in a natural manner to $(m+1)$-variate functions, as mentioned in section 2.4.1, which gives the desired functions $\widetilde{Q}_{h:n}(\boldsymbol{\theta}_h)$.

**Definition B.2.** *Let the new penalty function $Q_{\overline{m}:n}(\boldsymbol{\theta}_{\overline{m}})$ be given as follows,*

$$Q_{\overline{m}:n}(\boldsymbol{\theta}_{\overline{m}}) \coloneqq Q_{\overline{m}:n}(\boldsymbol{Y}_{\overline{m}:1}, \ldots, \boldsymbol{Y}_{\overline{m}:n}; \boldsymbol{\theta}_{\overline{m}}) \coloneqq \sum_{h=1}^{m} \widetilde{Q}_{h:n}(\boldsymbol{\theta}_h)\,, \tag{B.12a}$$

*where $\boldsymbol{\theta}_{\overline{m}}$ is the column vector obtained by stacking all the individual $\boldsymbol{\theta}_h$ on top of each other, i.e.*

$$\boldsymbol{\theta}_{\overline{m}} \coloneqq [\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_m']'\,. \tag{B.12b}$$

The $m$ components $\widetilde{Q}_{h:n}(\boldsymbol{\theta}_h)$ in the sum that defines $Q_{\overline{m}:n}(\boldsymbol{\theta}_{\overline{m}})$ have no common parameters, which implies that the optimisation of the parameters for the different summands can be performed independently. For a given sample from $\{\boldsymbol{Y}_{\overline{m}:t}\}_{t\in\mathbb{Z}}$ and for a given bandwidth $\boldsymbol{b}$, the optimal parameter vector $\widehat{\boldsymbol{\theta}}_{\overline{m}:n}$ for $Q_{\overline{m}:n}(\boldsymbol{\theta}_{\overline{m}})$ can thus be constructed by stacking on top of each other the parameter vectors that optimise the individual summands in eq. (B.12) — and these are the parameter vectors $\widehat{\boldsymbol{\theta}}_{h:n}$ that shows up for the $m$ bivariate cases in eq. (B.10). Since each $\widehat{\boldsymbol{\theta}}_{h:n}$ converge almost surely to $\boldsymbol{\theta}_{h:\boldsymbol{b}}$, it is clear that $\widehat{\boldsymbol{\theta}}_{\overline{m}:n}$ will converge

almost surely to $\boldsymbol{\theta}_{\bar{m}:\boldsymbol{b}}$, the vector obtained by stacking the $m$ vectors $\boldsymbol{\theta}_{h:\boldsymbol{b}}$ on top of each other.

The desired asymptotic result for the fixed $\boldsymbol{b}$ and fixed $m$ estimates $\widehat{f}_{\boldsymbol{v}}^{\,m}(\omega)$ can be obtained directly from the preceding observation and Theorem 1 in Tjøstheim and Hufthammer [2013], but that would not reveal how $m$ and $\boldsymbol{b}$ must behave in the general situation. The rest of this section will thus be used to verify items *(A1)* to *(A4)* from theorem B.1, which in essence only requires a minor adjustment of the bivariate discussion from appendix B.1.2, i.e. the discussion can start with the following Taylor-expansion of $Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}})$,

$$\begin{aligned} Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}}) = Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}:\boldsymbol{b}}) + [\boldsymbol{\theta}_{\bar{m}} - \boldsymbol{\theta}_{\bar{m}:\boldsymbol{b}}]' \boldsymbol{\nabla}_{\bar{m}} Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}:\boldsymbol{b}}) + \frac{1}{2} [\boldsymbol{\theta}_{\bar{m}} - \boldsymbol{\theta}_{\bar{m}:\boldsymbol{b}}]' V_{\bar{m}|\boldsymbol{b}:n} [\boldsymbol{\theta}_{\bar{m}} - \boldsymbol{\theta}_{\bar{m}:\boldsymbol{b}}] \\ + \frac{1}{2} [\boldsymbol{\theta}_{\bar{m}} - \boldsymbol{\theta}_{\bar{m}:\boldsymbol{b}}]' T_{\bar{m}|\boldsymbol{b}:n} [\boldsymbol{\theta}_{\bar{m}} - \boldsymbol{\theta}_{\bar{m}:\boldsymbol{b}}], \end{aligned} \tag{B.13}$$

where $\boldsymbol{\theta}_{\bar{m}:\boldsymbol{b}}$ represents the vector obtained by stacking on top of each other the $m$ individual population parameters $\boldsymbol{\theta}_{h:\boldsymbol{b}}$, where $\boldsymbol{\nabla}_{\bar{m}} := [\boldsymbol{\nabla}_1', \ldots, \boldsymbol{\nabla}_m']'$, and where the matrices $V_{\bar{m}|\boldsymbol{b}:n}$ and $T_{\bar{m}|\boldsymbol{b}:n}$ corresponds to the matrices $V_{h:\boldsymbol{b}:n}$ and $T_{h:\boldsymbol{b}:n}$ from eq. (B.9).

The following matrix-observations gives the foundation for the extension from the bivariate case to the multivariate case.

1. Keeping in mind how $\boldsymbol{\nabla}_{\bar{m}}$ is defined relative to $\boldsymbol{\nabla}_h$, and how $Q_{\bar{m}:n}$ is defined relative to $Q_{h:n}$, it is clear that $\boldsymbol{\nabla}_{\bar{m}} Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}:\boldsymbol{b}})$ is the vector obtained by stacking the $m$ vectors $\boldsymbol{\nabla}_h Q_{h:n}(\boldsymbol{\theta}_{h:\boldsymbol{b}})$ on top of each other.
2. The operator $\boldsymbol{\nabla}_{\bar{m}} \boldsymbol{\nabla}_{\bar{m}}'$ can be viewed as an $m \times m$ block-matrix, consisting of the $5 \times 5$ matrices $\boldsymbol{\nabla}_j \boldsymbol{\nabla}_k'$, $j, k = 1, \ldots, m$. Due to the definition of $Q_{\bar{m}:n}$, it is clear that the only operators $\boldsymbol{\nabla}_j \boldsymbol{\nabla}_k'$ that will return a nonzero result are those having $j = k$.
3. The preceding observation implies that $V_{\bar{m}|\boldsymbol{b}:n} = \bigoplus_{h=1}^{m} V_{h:\boldsymbol{b}:n}$, i.e. $V_{\bar{m}|\boldsymbol{b}:n}$ is the direct sum of the matrices $V_{h:\boldsymbol{b}:n}$ (the block diagonal matrix where the diagonal blocks equals $V_{h:\boldsymbol{b}:n}$, and all other blocks are zero, cf. e.g. Horn and Johnson [2012, p.30] for further details).
4. The same observation implies that $T_{\bar{m}|\boldsymbol{b}:n} = \bigoplus_{h=1}^{m} T_{h:\boldsymbol{b}:n}$

With these observations, and the details from the proof of Theorem 1 in Tjøstheim and Hufthammer [2013], it is straightforward to verify items *(A1)* to *(A3)* of theorem B.1, whereas item *(A4)* requires some more work.

**Lemma B.3** (Item *(A1)* of theorem B.1.)**.**
$n^{-1} \boldsymbol{\nabla}_{\bar{m}} Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}:\boldsymbol{b}}) \xrightarrow{a.s.} \mathbf{0}$

*Proof.* Since $\boldsymbol{\nabla}_{\bar{m}} Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}:\boldsymbol{b}})$ is the vector obtained by stacking the $m$ vectors $\boldsymbol{\nabla}_h Q_{h:n}(\boldsymbol{\theta}_{h:\boldsymbol{b}})$ on top of each other, and the proof of Theorem 1 in Tjøstheim and Hufthammer [2013] shows

that $n^{-1}\boldsymbol{\nabla}_h Q_{h:n}(\boldsymbol{\theta}_{h:\boldsymbol{b}})$ converges almost surely to $\mathbf{0}$, the same must necessarily be true for the combined vector $n^{-1}\boldsymbol{\nabla}_{\bar{m}} Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}:\boldsymbol{b}})$ too. □

**Lemma B.4** (Item *(A2)* of theorem B.1.)**.**
$n^{-1}V_{\bar{m}|\boldsymbol{b}:n} \xrightarrow{a.s.} V_{\bar{m}|\boldsymbol{b}}$*, where* $V_{\bar{m}|\boldsymbol{b}}$ *is a* $5m \times 5m$ *positive definite matrix.*

*Proof.* Since $V_{\bar{m}|\boldsymbol{b}:n}$ is the direct sum of the $m$ matrices $V_{h:\boldsymbol{b}:n}$, the behaviour of those will describe the behaviour of $V_{\bar{m}|\boldsymbol{b}:n}$. The proof of Theorem 1 in Tjøstheim and Hufthammer [2013] shows that the matrices $n^{-1}V_{h:\boldsymbol{b}:n}$ converges almost surely to positive definite matrices $V_{h:\boldsymbol{b}}$, and this implies that $n^{-1}V_{\bar{m}|\boldsymbol{b}:n}$ will converge almost surely to a block diagonal matrix $V_{\bar{m}|\boldsymbol{b}}$, defined as the direct sum of the matrices $V_{h:\boldsymbol{b}}$. Since the set of eigenvalues for a direct sum of matrices equals the union of the eigenvalues for its components, see Horn and Johnson [2012, p.30] for details, if follows that $V_{\bar{m}|\boldsymbol{b}:n}$ is positive definite since all the $V_{h:\boldsymbol{b}:n}$ are positive definite. □

**Lemma B.5** (Item *(A3)* of theorem B.1.)**.**
*For* $j, k = 1, \ldots, 5m$*,*

$$\lim_{n\to\infty} \sup_{\delta\to 0} (n\delta)^{-1} \left| T_{\bar{m}|\boldsymbol{b}:n\,jk} \right| < \infty \quad a.s., \tag{B.14}$$

*where* $T_{\bar{m}|\boldsymbol{b}:n\,jk}$ *is the* $(j,k)^{\text{th}}$ *component of* $T_{\bar{m}|\boldsymbol{b}:n}$*.*

*Proof.* $T_{\bar{m}|\boldsymbol{b}:n}$ is the direct sum of the $m$ matrices $T_{h:\boldsymbol{b}:n}$, so the required inequality is trivially satisfied for all entries $j$ and $k$ that gives an element outside of the diagonal-blocks. The proof of Theorem 1 in Tjøstheim and Hufthammer [2013] shows that the inequality is satisfied almost surely on each of the $m$ blocks $T_{h:\boldsymbol{b}:n}$, which implies that it holds for $T_{\bar{m}|\boldsymbol{b}:n}$ too. □

**Lemma B.6** (Item *(A4)* of theorem B.1.)**.**
$n^{-1/2}\boldsymbol{\nabla}_{\bar{m}} Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}:\boldsymbol{b}}) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, W_{\bar{m}|\boldsymbol{b}}\right)$

*Proof.* As done in the proof of Theorem 1 in Tjøstheim and Hufthammer [2013], the idea is to first prove asymptotic normality of each individual component of $\boldsymbol{\nabla}_{\bar{m}} Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}:\boldsymbol{b}})$ by Theorem 2.20(i) and Theorem 2.21(i) from Fan and Yao [2003, p. 74-75]. Then the Cramér-Wold Theorem (see e.g. Theorem 29.4 in Billingsley [2012]) will be used to conclude that the joint distribution of $\boldsymbol{\nabla}_{\bar{m}} Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}:\boldsymbol{b}})$ will be the joint distribution of these limiting components, and finally a simple observation based on moment-generating functions tells us that this limiting joint distribution is asymptotically normal.

Since $\boldsymbol{\nabla}_{\bar{m}} Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}:\boldsymbol{b}}) = [\boldsymbol{\nabla}_1 Q_{1:n}(\boldsymbol{\theta}_{h:\boldsymbol{b}})', \ldots, \boldsymbol{\nabla}_m Q_{m:n}(\boldsymbol{\theta}_{h:\boldsymbol{b}})']'$, its components can be indexed by pairs $[h, i]$, $h = 1, \ldots, m$ and $i = 1, \ldots, 5$. From eq. (B.8) it is clear that the $[h, i]$-component of the vector can be written as

$$(\boldsymbol{\nabla}_{\bar{m}} Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}:\boldsymbol{b}}))_{[h,i]} = -\sum_{t=1}^{n} X_{hi:t}, \tag{B.15}$$

where the random variable $X_{hi:t}$ is defined as

$$X_{hi:t} \coloneqq K_{h:\boldsymbol{b}}(\boldsymbol{Y}_{h:t} - \boldsymbol{v})\, u_{hi}(\boldsymbol{Y}_{h:t}; \boldsymbol{\theta}_{h:\boldsymbol{b}}) - \int_{\mathbb{R}^2} K_{h:\boldsymbol{b}}(\boldsymbol{y}_h - \boldsymbol{v})\, u_{hi}(\boldsymbol{y}_h; \boldsymbol{\theta}_{h:\boldsymbol{b}})\, \psi(\boldsymbol{y}_h; \boldsymbol{\theta}_h)\, \mathrm{d}\boldsymbol{y}_h, \tag{B.16}$$

and where $u_{hi}$ refers to the $i^{\text{th}}$ component of the $h^{\text{th}}$ score function $\boldsymbol{u}_h$.

The required $\alpha$-mixing property (and thus ergodicity) are inherited from the original univariate time series $Y_t$ to $X_{hi:t}$ (see eq. (C.36) for details), and the connection with $L^{\nu}$-theory observed in eq. (C.41) gives $\mathrm{E}[|X_{hi:t}|^{\nu}] < \infty$. Finally, since $\boldsymbol{\theta}_{h:\boldsymbol{b}}$ is the population value parameter that minimise eq. (B.5), it follows that $\mathrm{E}[X_{hi:t}] = 0$. These observations show that $X_{hi:t}$ satisfies the requirements needed in order to apply Theorem 2.20(i) and Theorem 2.21(i) from Fan and Yao [2003, p. 74-75], i.e. for $S_{hi|n} \coloneqq \sum_{t=1}^{n} X_{hi:t}$, Theorem 2.20(i) gives the asymptotic result

$$n^{-1} S_{hi|n} \longrightarrow \sigma^2 \coloneqq \gamma_0 + 2 \sum_{\ell \geq 1} \gamma_\ell, \tag{B.17}$$

with $\gamma_\ell$ being the $\ell^{\text{th}}$ autocovariance of the series $\{X_{hi:t}\}_{t \in \mathbb{Z}}$. From Theorem 2.21(i) it now follows that there is a component-wise asymptotic normality, i.e.

$$n^{-1/2} S_{hi|n} \xrightarrow{d} \mathrm{N}(0, \sigma^2)\,. \tag{B.18}$$

In order to apply the Cramér-Wold device, all possible linear combinations of the components in $\boldsymbol{\nabla}_{\bar{m}} Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}:\boldsymbol{b}})$ must be considered. Such general sums can be represented as $S_n(\boldsymbol{a}) \coloneqq \boldsymbol{a}'\, \boldsymbol{\nabla}_{\bar{m}} Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}:\boldsymbol{b}})$, where $\boldsymbol{a} \in \mathbb{R}^{5 \times m}$. This can be rewritten, by 'taking the sum outside of the vector $\boldsymbol{\nabla}_{\bar{m}} Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}:\boldsymbol{b}})$', as

$$S_n(\boldsymbol{a}) = \sum_{t=1}^{n} X_t(\boldsymbol{a})\,, \tag{B.19}$$

where $X_t(\boldsymbol{a}) = \boldsymbol{a}' \boldsymbol{X}_t$, with the vector $\boldsymbol{X}_t$ obtained by stacking all the components $X_{hi:t}$ on top of each other, i.e. $\boldsymbol{X}_t = [X_{11:t}, \ldots, X_{m5:t}]'$.

By construction, $\mathrm{E}[X_t(\boldsymbol{a})] = 0$, the required $\alpha$-mixing are inherited from the original time series $\{Y_t\}$ (see eq. (C.36)), and lemma C.8 ensures that the property $\mathrm{E}[|X_t(\boldsymbol{a})|^\nu] < \infty$ holds true. That is, $X_t(\boldsymbol{a})$ does also satisfy the requirements stated in Theorem 2.20(i) and Theorem 2.21(i), which gives the following asymptotic results;

$$n^{-1}S_n(\boldsymbol{a}) \longrightarrow \sigma^2(\boldsymbol{a}) \coloneqq \gamma_0(\boldsymbol{a}) + 2\sum_{\ell\geq 1}\gamma_\ell(\boldsymbol{a}) \tag{B.20}$$

$$n^{-1/2}S_n(\boldsymbol{a}) \xrightarrow{d} \mathrm{N}(0, \sigma^2(\boldsymbol{a}))\,, \tag{B.21}$$

where the autocovariances $\gamma_\ell(\boldsymbol{a})$ now are with respect to the time series $X_t(\boldsymbol{a}) = \boldsymbol{a}'\boldsymbol{X}_t$.

Since $\gamma_0(\boldsymbol{a}) = \mathrm{Var}(\boldsymbol{a}'\boldsymbol{X}_t) = \boldsymbol{a}'\,\mathrm{Var}(\boldsymbol{X}_t)\,\boldsymbol{a}$ and $\gamma_\ell(\boldsymbol{a}) = \mathrm{Cov}\big(\boldsymbol{a}'\boldsymbol{X}_{t+\ell}, \boldsymbol{a}'\boldsymbol{X}_t\big) = \boldsymbol{a}'\,\mathrm{Cov}\big(\boldsymbol{X}_{t+\ell}, \boldsymbol{X}_t\big)\,\boldsymbol{a}$, it follows that we can write $\sigma^2(\boldsymbol{a}) = \boldsymbol{a}'W_{\bar{m}|\boldsymbol{b}}\boldsymbol{a}$, with $W_{\bar{m}|\boldsymbol{b}}$ being the matrix obtained in the obvious manner by factorising out $\boldsymbol{a}'$ and $\boldsymbol{a}$ from the sum of autocovariances, i.e.

$$W_{\bar{m}|\boldsymbol{b}} \coloneqq \mathrm{Var}(\boldsymbol{X}_t) + 2\sum_{\ell\geq 1}\mathrm{Cov}\big(\boldsymbol{X}_{t+\ell}, \boldsymbol{X}_t\big) \tag{B.22}$$

$$= \mathrm{E}[\boldsymbol{X}_t\boldsymbol{X}_t'] + 2\sum_{\ell\geq 1}\mathrm{E}\big[\boldsymbol{X}_{t+\ell}\boldsymbol{X}_t'\big]\,, \tag{B.23}$$

where the second equality follows since $\mathrm{E}[\boldsymbol{X}_t] = \boldsymbol{0}$.

The Cramér-Wold device now gives the required conclusion, $n^{-1/2}\boldsymbol{\nabla}_{\bar{m}}Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}:\boldsymbol{b}}) \xrightarrow{d} \mathrm{N}\big(\boldsymbol{0}, W_{\bar{m}|\boldsymbol{b}}\big)$. □

Lemmas B.3 to B.6 shows that the penalty function $Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}})$ (for fixed $m$ and fixed $\boldsymbol{b}$) satisfies the four requirements given in items *(A1)* to *(A4)* of theorem B.1, and this implies that the following asymptotic results holds in this particular case

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_{\bar{m}:n} - \boldsymbol{\theta}_{\bar{m}:\boldsymbol{b}}\right) \xrightarrow{d} \mathrm{N}\big(\boldsymbol{0}, V_{\bar{m}|\boldsymbol{b}}^{-1}W_{\bar{m}|\boldsymbol{b}}V_{\bar{m}|\boldsymbol{b}}^{-1}\big)\,. \tag{B.24}$$

The hard task to deal with in the general situation, when $m \to \infty$ and $\boldsymbol{b} \to \boldsymbol{0}^+$ as $n \to \infty$, is the asymptotic behaviour of $n^{-1/2}\boldsymbol{\nabla}_{\bar{m}}Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}:\boldsymbol{b}})$. This will be treated in appendix B.2.

## B.2 The A4-requirement in the general case

The verification of the three first requirements of the Klimko-Nelson approach does work as before when '$m \to \infty$ and $\boldsymbol{b} \to \boldsymbol{0}^+$ when $n \to \infty$', whereas the asymptotic normality in the

fourth requirement demands a more detailed investigation. Appendix B.2.1 will introduce some new building blocks to be used in the investigation of the asymptotic properties, which will be developed in appendices B.2.2 and B.2.3. Some technical details that only depend upon the kernel function and the score functions have been collected in appendix C.4.

### B.2.1 The final building blocks

The bivariate processes $\boldsymbol{Y}_{h:t}$ from definition 2.7 will now be used to construct new random variables, that culminates in a random variable $\mathfrak{Q}_{\bar{m}}^{n}$ which has the same limiting distribution[6] $\sqrt{b_1 b_2}\boldsymbol{\nabla}_{\bar{m}} Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}:\boldsymbol{b}})$. Looking upon eq. (B.8), it is clear that everything depends upon the three functions $\psi(\boldsymbol{y}_h;\boldsymbol{\theta}_h)$, $\boldsymbol{u}_h(\boldsymbol{y}_h;\boldsymbol{\theta}_h)$ and $K_{h:\boldsymbol{b}}(\boldsymbol{y}_h - \boldsymbol{v})$.

**Definition B.7.** *For $\psi(\boldsymbol{y}_h;\boldsymbol{\theta}_h)$ the local Gaussian density used when approximating $g_h(\boldsymbol{y}_h)$ at the point $\boldsymbol{v} = (v_1, v_2)$, define for all $h \in \mathbb{N}$ and $q \in \{1, \dots, 5\}$*

*(a) With $\boldsymbol{\theta}_{h:\boldsymbol{b}}$ the population value that minimises the penalty function $q_{h:\boldsymbol{b}}$ from eq. (B.4), let*

$$u_{hq:\boldsymbol{b}}(\boldsymbol{w}) := \frac{\partial}{\partial \theta_{h:q}} \log\left(\psi(\boldsymbol{y}_h;\boldsymbol{\theta}_h)\right)\Big|_{(\boldsymbol{y}_h;\boldsymbol{\theta}_h)=(\boldsymbol{w};\boldsymbol{\theta}_{h:\boldsymbol{b}})}. \tag{B.25}$$

*(b) For $L \geq 0$, define the following lower and upper truncated versions of $u_{hq:\boldsymbol{b}}(\boldsymbol{w})$,*

$$u_{hq:\boldsymbol{b}}(\boldsymbol{w})^{\leq L} := u_{hq:\boldsymbol{b}}(\boldsymbol{w}) \cdot \mathbb{1}\big\{\big|u_{hq:\boldsymbol{b}}(\boldsymbol{w})\big| \leq L\big\}, \tag{B.26a}$$

$$u_{hq:\boldsymbol{b}}(\boldsymbol{w})^{>L} := u_{hq:\boldsymbol{b}}(\boldsymbol{w}) \cdot \mathbb{1}\big\{\big|u_{hq:\boldsymbol{b}}(\boldsymbol{w})\big| > L\big\}. \tag{B.26b}$$

*Obviously; $u_{hq:\boldsymbol{b}}(\boldsymbol{w}) = u_{hq:\boldsymbol{b}}(\boldsymbol{w})^{\leq L} + u_{hq:\boldsymbol{b}}(\boldsymbol{w})^{>L}$ and $u_{hq:\boldsymbol{b}}(\boldsymbol{w})^{\leq L} \cdot u_{hq:\boldsymbol{b}}(\boldsymbol{w})^{>L} = 0$.*

*(c) Let $u_{hq}(\boldsymbol{w})$ be as in item (a), with the difference that the limit $\boldsymbol{b} \to \boldsymbol{0}^+$ of the parameters $\boldsymbol{\theta}_{h:\boldsymbol{b}}$ are used in the definition.[7] Let $u_{hq}(\boldsymbol{w})^{\leq L}$ and $u_{hq}(\boldsymbol{w})^{>L}$ be the truncated versions of $u_{hq}(\boldsymbol{w})$.*

The following simple observations will be useful later on.

**Lemma B.8.** *For the point $\boldsymbol{v}$, the following holds for the functions introduced in definition B.7.*

[6] Due to the presence of the kernel function $K_{h:\boldsymbol{b}}(\boldsymbol{w})$, the fourth requirement of the Klimko-Nelson approach will (when $\boldsymbol{b} \to \boldsymbol{0}^+$) require that the scaling factor $n^{-1/2}$ is adjusted with $(b_1 b_2)^{1/2}$, and this scaling must thus also be included in the discussion in the present approach.

[7] The limit of the parameters $\boldsymbol{\theta}_{h:\boldsymbol{b}}$ will exist under assumptions that implies that the four requirements of the Klimko-Nelson approach are satisfied, cf. Tjøstheim and Hufthammer [2013] for details.

*(a)* $\sup_{hq} \left|u_{hq:\boldsymbol{b}}(\boldsymbol{v})\right| < \infty$ *and* $\sup_{hq} \left|u_{hq}(\boldsymbol{v})\right| < \infty$.
*(b) When* $L$ *is large enough,* $u_{hq:\boldsymbol{b}}(\boldsymbol{v})^{\leq L} = u_{hq:\boldsymbol{b}}(\boldsymbol{v})$ *and* $u_{hq}(\boldsymbol{v})^{\leq L} = u_{hq}(\boldsymbol{v})$.

*Proof.* By definition, the functions $u_{hq:\boldsymbol{b}}(\boldsymbol{w})$ and $u_{hq}(\boldsymbol{w})$ will all be bivariate polynomials of order two (in the variables $w_1$ and $w_2$), which implies that they are well defined for any point $\boldsymbol{v}$. Since the parameters in these polynomials originates from a local Gaussian approximation of $g_h(\boldsymbol{y}_h)$ at the point $\boldsymbol{v}$, and since assumption 2.1(b) ensures that the bivariate densities $g_h(\boldsymbol{y}_h)$ will approach the product of the marginal densities when $h \to \infty$, it follows that the estimated parameters must stabilise when $h$ becomes large. This rules out the possibility that any of the parameters can grow to infinitely large values, which implies that the supremums in item (a) are finite. Item (b) follows as a direct consequence of this, the statement holds true for any threshold value $L$ that is larger than the supremums given in item (a). □

The bivariate kernel to be used in the present approach will be the same as the one used in Tjøstheim and Hufthammer [2013], i.e. it will be the product kernel based on two standard normal kernels. The following definition enables a more general approach to be used in the theoretical investigation,[8] while capturing the desirable properties that will be satisfied for the product normal kernel.

**Definition B.9.** *From a bivariate, non-negative, and bounded kernel function* $K(\boldsymbol{w})$*, that satisfies*

$$\int_{\mathbb{R}^2} K(w_1, w_2)\, \mathrm{d}w_1 \mathrm{d}w_2 = 1, \tag{B.27a}$$

$$\mathcal{K}_{1:k}(w_2) \coloneqq \int_{\mathbb{R}^1} K(w_1, w_2)\, w_1^k\, \mathrm{d}w_1 \qquad \text{is bounded for } k \in \{0, 1, 2\}, \tag{B.27b}$$

$$\mathcal{K}_{2:\ell}(w_1) \coloneqq \int_{\mathbb{R}^1} K(w_1, w_2)\, w_2^\ell\, \mathrm{d}w_2 \qquad \text{is bounded for } \ell \in \{0, 1, 2\}, \tag{B.27c}$$

$$\int_{\mathbb{R}^2} K(w_1, w_2)\, |w_1^k w_2^\ell|\, \mathrm{d}w_1 \mathrm{d}w_2 < \infty, \qquad k, \ell \geq 0 \text{ and } k + \ell \leq 2 \cdot \lceil \nu \rceil\,, \tag{B.27d}$$

*where* $\nu > 2$ *is from assumption 2.1(b) (and* $\lceil \cdot \rceil$ *is the ceiling function), define*

$$K_{h:\boldsymbol{b}}(\boldsymbol{y}_h - \boldsymbol{v}) \coloneqq \frac{1}{b_1 b_2} K\left(\frac{y_h - v_1}{b_1}, \frac{y_0 - v_2}{b_2}\right). \tag{B.28}$$

[8] Differences in the computational cost implies that the product normal kernel is used for practical purposes.

It turns out, see appendix C.4 for details, that the asymptotic results needed later on mainly depends upon the properties of the kernel $K(\boldsymbol{w})$ and the components $u_{hq:\boldsymbol{b}}(\boldsymbol{w})$ of the score functions.

Some vector and matrix notation is needed in order to make the expressions later on more tractable.

**Definition B.10.** *With $g_h(\boldsymbol{y}_h)$, $u_{hq:\boldsymbol{b}}(\boldsymbol{w})$ and $K(\boldsymbol{w})$ as given in definitions 2.7, B.7 and B.9, let $\mathfrak{U}_{h:\boldsymbol{b}} \coloneqq \left[u_{h1:\boldsymbol{b}}(\boldsymbol{v}), \ldots, u_{hp:\boldsymbol{b}}(\boldsymbol{v})\right]'$, and define the following matrices.*

$$W_{h:\boldsymbol{b}} \coloneqq \mathfrak{U}_{h:\boldsymbol{b}}\mathfrak{U}'_{h:\boldsymbol{b}} \cdot g_h(\boldsymbol{v}) \int_{\mathbb{R}^2} K(\boldsymbol{w})^2 \,\mathrm{d}\boldsymbol{w}, \tag{B.29a}$$

$$W_{\overline{m}|b} \coloneqq \bigoplus_{h=1}^{m} W_{h:\boldsymbol{b}}. \tag{B.29b}$$

*Matrices $W_h$ and $W_{\overline{m}}$ can be defined in a similar manner, using the $\boldsymbol{b} \to \mathbf{0}^+$ versions $u_{hq}(\boldsymbol{w})$ from definition B.7(c). Note that $W_{h:\boldsymbol{b}}$ and $W_h$ will have rank one, whereas $W_{\overline{m}:\boldsymbol{b}}$ and $W_{\overline{m}}$ will have rank $m$. Furthermore, note that if $\boldsymbol{a}_h \in \mathbb{R}^5$ and $\boldsymbol{a}_{\overline{m}} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m]'$, then $\boldsymbol{a}'_{\overline{m}} W_{\overline{m}:\boldsymbol{b}} \boldsymbol{a}_{\overline{m}} = \sum_{h=1}^{m} \boldsymbol{a}'_h W_{h:\boldsymbol{b}} \boldsymbol{a}_h$.*

The time is due for the introduction of the random variables.

**Definition B.11.** *Based on $\boldsymbol{Y}_{h:t}$, $u_{hq:\boldsymbol{b}}(\boldsymbol{w})$ and $K_{h:\boldsymbol{b}}(\boldsymbol{y}_h - \boldsymbol{v})$ from definitions 2.7, B.7 and B.9, define new bivariate random variables as follows,*

$$X^n_{hq:t}(\boldsymbol{v}) \coloneqq \sqrt{b_1 b_2} K_{h:\boldsymbol{b}}(\boldsymbol{Y}_{h:t} - \boldsymbol{v})\, u_{hq:\boldsymbol{b}}(\boldsymbol{Y}_{h:t}), \tag{B.30a}$$

$$X^{n|\leq L}_{hq:t}(\boldsymbol{v}) \coloneqq \sqrt{b_1 b_2} K_{h:\boldsymbol{b}}(\boldsymbol{Y}_{h:t} - \boldsymbol{v})\, u_{hq:\boldsymbol{b}}(\boldsymbol{Y}_{h:t})^{\leq L}, \tag{B.30b}$$

$$X^{n|>L}_{hq:t}(\boldsymbol{v}) \coloneqq \sqrt{b_1 b_2} K_{h:\boldsymbol{b}}(\boldsymbol{Y}_{h:t} - \boldsymbol{v})\, u_{hq:\boldsymbol{b}}(\boldsymbol{Y}_{h:t})^{>L}. \tag{B.30c}$$

*Obviously; $X^n_{hq:t}(\boldsymbol{v}) = X^{n|\leq L}_{hq:t}(\boldsymbol{v}) + X^{n|>L}_{hq:t}(\boldsymbol{v})$ and $X^{n|\leq L}_{hq:t}(\boldsymbol{v}) \cdot X^{n|>L}_{hq:t}(\boldsymbol{v}) = 0$.*

Since the point $\boldsymbol{v}$ will be fixed for the remainder of this discussion, $\boldsymbol{v}$ will be suppressed and only $X^n_{hq:t}$ will be used when referring to eq. (B.30a), and $\boldsymbol{v}$ will also be suppressed for the new random variables derived from $X^n_{hq:t}$.

Note: A comparison of $X^n_{hq:t}$ against the components occurring in the expression for $\boldsymbol{\nabla}_h Q_{h:n}(\boldsymbol{\theta}_h)$, see eq. (B.8), implies that the following adjusted variable should be included,

$$\widetilde{X}^n_{hq:t} \coloneqq X^n_{hq:t} - \sqrt{b_1 b_2} \int_{\mathbb{R}^2} K_{h:\boldsymbol{b}}(\boldsymbol{y}_h - \boldsymbol{v})\, u_{hq:\boldsymbol{b}}(\boldsymbol{y}_h)\, \psi(\boldsymbol{y}_h; \boldsymbol{\theta}_h)\, \mathrm{d}\boldsymbol{y}_h, \tag{B.31}$$

but the arguments later on will use a mean adjusted approach similar to the one used in Masry and Tjøstheim [1995], see the definitions of $Z^n_{hq:t}$ and $\mathcal{Q}^n_{hq}$ below, and the only place $\widetilde{X}^n_{hq:t}$ is needed is in the proof of lemma B.14.

**Definition B.12.** *Based on the bivariate random variables $X^n_{hq:t}$ from definition B.11 define the following bivariate and $(m+1)$-variate random variables,*

$$Z^n_{hq:t} \coloneqq X^n_{hq:t} - \mathrm{E}\big[X^n_{hq:t}\big]\,, \tag{B.32a}$$

$$\mathcal{Q}^n_{hq} \coloneqq \sum_{t=1}^{n} Z^n_{hq:t}. \tag{B.32b}$$

*Similarly, $Z^{n|\geq L}_{hq:t}$, $Z^{n|<L}_{hq:t}$, $\mathcal{Q}^{n|\geq L}_{hq}$ and $\mathcal{Q}^{n|<L}_{hq}$ can be defined in the natural manner, with the obvious connections $Z^n_{hq:t} = Z^{n|\geq L}_{hq:t} + Z^{n|<L}_{hq:t}$, $Z^{n|\geq L}_{hq:t} \cdot Z^{n|<L}_{hq:t} = 0$, and $\mathcal{Q}^n_{hq} = \mathcal{Q}^{n|\geq L}_{hq} + \mathcal{Q}^{n|<L}_{hq}$ holding for all $L$. Moreover: $\mathrm{Cov}\big(Z^n_{hq:i}, Z^n_{j:k}\big) = \mathrm{E}\big[Z^n_{hq:i} \cdot Z^n_{j:k}\big] = \mathrm{Cov}\big(X^n_{hq:i}, X^n_{jr:k}\big)$.*

The last batch of random variables can now be introduced.

**Definition B.13.** *Based upon the bivariate $Z^n_{hq:t}$ from definition B.12, and for $\boldsymbol{a} \coloneqq \boldsymbol{a}_{\overline{m}} \in \mathbb{R}^{5\times m}$, define the following $(m+1)$-variate random variables,*

$$Z^n_{\overline{m}:t}(\boldsymbol{a}) \coloneqq \sum_{h=1}^{m}\sum_{q=1}^{5} a_{hq} Z^n_{hq:t} = \boldsymbol{a}'\boldsymbol{Z}^n_{\overline{m}:t}, \tag{B.33a}$$

$$\mathcal{Q}^n_{\overline{m}}(\boldsymbol{a}) \coloneqq \sum_{h=1}^{m}\sum_{q=1}^{5} a_{hq} \mathcal{Q}^n_{hq} = \boldsymbol{a}'\boldsymbol{\mathcal{Q}}^n_{\overline{m}}, \tag{B.33b}$$

*where $\boldsymbol{Z}^n_{\overline{m}:t}$ and $\boldsymbol{\mathcal{Q}}^n_{\overline{m}}$ are defined in the obvious manner.*

**Lemma B.14.** *$\boldsymbol{\mathcal{Q}}^n_{\overline{m}}$ and $\sqrt{b_1 b_2}\,\boldsymbol{\nabla}_{\overline{m}} Q_{\overline{m}:n}(\boldsymbol{\theta}_{\overline{m}:\boldsymbol{b}})$ share the same limiting distribution.*

*Proof.* The only difference between $\boldsymbol{\mathcal{Q}}^n_{\overline{m}}$ and $\sqrt{b_1 b_2}\,\boldsymbol{\nabla}_{\overline{m}} Q_{\overline{m}:n}(\boldsymbol{\theta}_{\overline{m}:\boldsymbol{b}})$ is that the first use $Z^n_{hq:t}$ where the second use $\widetilde{X}^n_{hq:t}$. The difference between these components are

$$Z^n_{hq:t} - \widetilde{X}^n_{hq:t} = \sqrt{b_1 b_2} \cdot \int_{\mathbb{R}^2} K_{h:\boldsymbol{b}}(\boldsymbol{y}_h - \boldsymbol{v})\, u_{hq:\boldsymbol{b}}(\boldsymbol{y}_h) \left\{ g_h(\boldsymbol{y}_h) - \psi(\boldsymbol{y}_h; \boldsymbol{\theta}_h) \right\} \mathrm{d}\boldsymbol{y}_{\boldsymbol{h}}, \tag{B.34}$$

and this difference will not only approach zero but in fact be identical to zero when the bandwidth $\boldsymbol{b}$ is smaller than $\boldsymbol{b}_0$, since the population value $\boldsymbol{\theta}_{h:\boldsymbol{b}}$ in that case satisfies eq. (B.5). The result now follows from Billingsley [2012, Th. 25.4]. □

The purpose of the new random variables introduced in definitions B.11 to B.13 is to find under which conditions the fourth requirement of the Klimko-Nelson approach is satisfied in the general situation where $m \to \infty$ and $\boldsymbol{b} \to \mathbf{0}^{+}$ when $n \to \infty$.

The part that does require some effort to investigate is the fourth requirement of theorem B.1, which (using the notation introduced here) means that it is necessary to verify that $n^{-1/2}\,\mathfrak{Q}_{\overline{m}}^{n}$ approaches a normal distribution when $\boldsymbol{b}$ goes to zero when $n$ and $m$ are 'large enough'. The proof will be presented in a step by step manner, that builds upon the asymptotic behaviour of $\mathrm{E}\big[X_{hq:i}^{n} \cdot X_{jr:k}^{n}\big]$. The computation of this expectation will (depending on the indices $h$, $i$, $j$ and $k$) either require a bivariate, trivariate or tetravariate integral.

| Combinations | $\boldsymbol{v}$ | $\boldsymbol{b}$ | $\boldsymbol{Y}_{h:i}$ | $\boldsymbol{Y}_{j:k}$ |
|---|---|---|---|---|
| First argument of $K_{h:\boldsymbol{b}}$ | $v_1$ | $b_1$ | $Y_{h+i}$ | $Y_{j+k}$ |
| Second argument of $K_{h:\boldsymbol{b}}$ | $v_2$ | $b_2$ | $Y_i$ | $Y_k$ |

Table 1: Factors deciding bivariate, trivariate or tetravariate.

Table 1 lists the combinations that must be taken into account when computing $\mathrm{E}\big[X_{hq:i}^{n} \cdot X_{jr:k}^{n}\big]$, i.e. the presence of $\boldsymbol{v}$ and $\boldsymbol{b}$ and the dependence on $Y_t$ in the kernel functions — and it is evident from this table that the amount of overlap in the indexing set $\{i, h+i, k, j+k\}$ will decide if the resulting integral turns out to be bi-, tri- or tetravariate. Note that eq. (2.17) of algorithm 2.5(c) implies that only positive indices are required, so the bivariate case can thus only occur when $i = k$ and $h = j$. It will be seen later on that these bivariate components are the only ones that adds non-negligible contributions to the asymptotic behaviour.

### B.2.2 The asymptotic results — basic part

The analysis of the asymptotic properties of $X_{hq:i}^{n}$, from definition B.11, would be quite simple if either the kernel function $K(\boldsymbol{w})$ or the score-function components $u_{hq:\boldsymbol{b}}(\boldsymbol{w})$ had bounded support, since the finiteness requirements of assumption 2.1(g) then would follow directly from lemma C.6, and the proof of lemma B.15 would be rather trivial. However, in the present analysis, $K(\boldsymbol{w})$ and $u_{hq:\boldsymbol{b}}(\boldsymbol{w})$ both have $\mathbb{R}^2$ as their support, which implies that extra care must be taken when working with the densities under consideration.

**Lemma B.15.** *When $Y_t$ satisfies assumption 2.1, and $u_{hq:\boldsymbol{b}}(\boldsymbol{w})$ and $K(\boldsymbol{w})$ are as given in definitions B.7 and B.9, then the random variables $X_{hq:t}^{n}$ from definition B.11 satisfies*

*(a)* $\mathrm{E}\big[X_{hq:i}^{n}\big] = O\big(\sqrt{b_1 b_2}\big)$.

*(b)* $\mathrm{E}\big[\big|X^n_{hq:i}\big|^\nu\big]^{1/\nu} = O\big(|b_1 b_2|^{(2-\nu)/2\nu}\big)$.

*(c)* $\mathrm{E}\big[X^n_{hq:i} \cdot X^n_{jr:k}\big] = \begin{cases} u_{hq:\boldsymbol{b}}(\boldsymbol{v})\, u_{hr:\boldsymbol{b}}(\boldsymbol{v})\, g_h(\boldsymbol{v}) \int_{\mathbb{R}^2} K(\boldsymbol{w})^2 \,\mathrm{d}\boldsymbol{w} + O(b_1 \vee b_2) & \text{when bivariate,} \\ O(b_1 \wedge b_2) & \text{when trivariate,} \\ O(b_1 b_2) & \text{when tetravariate,} \end{cases}$

*where bivariate, trivariate and tetravariate refers to how many different $Y_t$ the four indices $h$, $i$, $j$ and $k$ gives, cf. table 1 for details.*

*Proof.* The expectations in items (a) to (c) are all finite due to assumption 2.1(g) and they do in addition correspond to integrals whose integrands are of the form $\mathcal{V} \cdot g$, where $g$ is a density function and $\mathcal{V}$ is an integrand of the type discussed in items (a) to (c) of lemma C.6, i.e. $\mathcal{V}$ collects everything that only depends on the functions $u_{hq:\boldsymbol{b}}(\boldsymbol{w})$ and $K(\boldsymbol{w})$. The substitutions used in the proof of lemma C.6 can be applied to the different cases under investigation, and it follows that these substitutions will create new integrals with the desired function of $b_1$ and $b_2$ as a scaling factor. This proves items (a) and (b) and it also takes care of the trivariate and tetravariate cases of item (c).

Equation (2.22) from assumption 2.1(d) is needed for the bivariate case of item (c), i.e. the Taylor expansion of $g_h(\boldsymbol{y}_h)$ around the point $\boldsymbol{v}$ allows the integral of interest to be written as the sum of the following three integrals:

$$\mathcal{J}_1 \coloneqq \int_{\mathbb{R}^2} \mathcal{V}(\boldsymbol{y}_h) \cdot g_h(\boldsymbol{v}) \,\mathrm{d}\boldsymbol{y}_h, \tag{B.35a}$$

$$\mathcal{J}_2 \coloneqq \int_{\mathbb{R}^2} \mathcal{V}(\boldsymbol{y}_h) \cdot \left(\boldsymbol{\mathfrak{g}}_h(\boldsymbol{v})' \left[\boldsymbol{y}_h - \boldsymbol{v}\right]\right) \mathrm{d}\boldsymbol{y}_h, \tag{B.35b}$$

$$\mathcal{J}_3 \coloneqq \int_{\mathbb{R}^2} \mathcal{V}(\boldsymbol{y}_h) \cdot \left(\boldsymbol{\mathfrak{R}}_h(\boldsymbol{y}_h)' \left[\boldsymbol{y}_h - \boldsymbol{v}\right]\right) \mathrm{d}\boldsymbol{y}_h. \tag{B.35c}$$

The bivariate case of lemma C.6(c) shows that the term $\mathcal{J}_1$ gives the desired result, so it remains to prove that the terms $\mathcal{J}_2$ and $\mathcal{J}_3$ are $O(b_1 \vee b_2)$. For this investigation, the substitution $w_1 = (y_h - v_1)/b_1$ and $w_2 = (y_0 - v_2)/b_2$ must be applied, which in particular replaces the vector $[\boldsymbol{y}_h - \boldsymbol{v}]$ with the vector $[b_1 w_1, b_2 w_2]'$. In order to compactify the notation, let $a_1$ and $a_2$ denote the two components of $\boldsymbol{\mathfrak{g}}_h(\boldsymbol{v})$, let $\mathcal{W}$ be the substituted version of $\mathcal{V}$, let $\mathfrak{R}_{h1}$ and $\mathfrak{R}_{h2}$ be the two components of the remainder function and finally let $\mathfrak{T}_{h1}$ and $\mathfrak{T}_{h2}$ be the substituted versions of $\mathfrak{R}_{h1}\mathcal{W}$ and $\mathfrak{R}_{h2}\mathcal{W}$.

With this notation, the substitution used upon $\mathcal{J}_2$ gives

$$\mathcal{J}_2 = a_1 b_1 \int_{\mathbb{R}^2} w_1 \cdot \mathcal{W}(\boldsymbol{w}) \,\mathrm{d}\boldsymbol{w} + a_2 b_2 \int_{\mathbb{R}^2} w_2 \cdot \mathcal{W}(\boldsymbol{w}) \,\mathrm{d}\boldsymbol{w}, \tag{B.36}$$

whose integrands include an extra factor of $w_1$ or $w_2$ compared to the integrands encountered in the proof of lemma C.6. This is however no problem, since lemma C.5(b) implies that the finiteness conclusion still holds true in these cases, which implies that $\mathcal{J}_2$ is $O(b_1 \vee b_2)$

Since assumption 2.1(g) ensures that the sum of the three integrals $\mathcal{J}_1$, $\mathcal{J}_2$ and $\mathcal{J}_3$ is finite, and the above discussion shows that the two first of them are finite, it follows that $\mathcal{J}_3$ also is finite. An inspection of $\mathcal{J}_3$ after substitution, i.e.

$$\mathcal{J}_3 = \int_{\mathbb{R}^2} [b_1 w_1 \cdot \mathfrak{T}_{h1}(\boldsymbol{y}(\boldsymbol{w})) + b_2 w_2 \cdot \mathfrak{T}_{h2}(\boldsymbol{y}(\boldsymbol{w}))] \, \mathrm{d}\boldsymbol{w}, \tag{B.37}$$

then reveal that the maximum of $b_1$ and $b_2$ can be factorised out of the integrand. This implies that $\mathcal{J}_3$ is $O(b_1 \vee b_2)$, and thus concludes the proof of lemma B.15 □

The following corollary is handy when the covariance is the target of interest.

**Corollary B.16.** *When $Y_t$ satisfies assumption 2.1, and $u_{hq:\boldsymbol{b}}(\boldsymbol{w})$ and $K(\boldsymbol{w})$ are as given in definitions B.7 and B.9, then the random variables $X^n_{hq:t}$ from definition B.11 satisfies*

$$\mathrm{Cov}\big(X^n_{hq:i}, X^n_{jr:k}\big) = \begin{cases} u_{hq:\boldsymbol{b}}(\boldsymbol{v})\, u_{hr:\boldsymbol{b}}(\boldsymbol{v})\, g_h(\boldsymbol{v}) \int_{\mathbb{R}^2} K(\boldsymbol{w})^2 \, \mathrm{d}\boldsymbol{w} + O(b_1 \vee b_2) & \text{when bivariate,} \\ O(b_1 \wedge b_2) & \text{when trivariate,} \\ O(b_1 b_2) & \text{when tetravariate.} \end{cases} \tag{B.38}$$

*Proof.* Since $\mathrm{Cov}\big(X^n_{hq:i}, X^n_{jr:k}\big) = \mathrm{E}\big[X^n_{hq:i} \cdot X^n_{jr:k}\big] - \mathrm{E}\big[X^n_{hq:i}\big] \cdot \mathrm{E}\big[X^n_{jr:k}\big]$, the result follows immediately from an inspection of items (a) and (c) of lemma B.15. □

The next corollary is needed in the proof of lemma B.18.

**Corollary B.17.** *When $Y_t$ satisfies assumption 2.1, and $u_{hq:\boldsymbol{b}}(\boldsymbol{w})$ and $K(\boldsymbol{w})$ are as given in definitions B.7 and B.9, then the random variables $Z^n_{hq:t}$ and $Z^n_{\bar{m}:t}(\boldsymbol{a})$ from definition B.12 satisfies*

*(a)* $\mathrm{E}\big[\big|Z^n_{hq:t}\big|^\nu\big]^{1/\nu} = O\big(|b_1 b_2|^{(2-\nu)/2\nu}\big)$.

*(b)* $\mathrm{E}[|Z^n_{\bar{m}:t}(\boldsymbol{a})|^\nu]^{1/\nu} = O\big(m\,|b_1 b_2|^{(2-\nu)/2\nu}\big)$.

*Proof.* The connection between expectations and $L^\nu$-spaces discussed in appendix C.5, see eq. (C.41), can be applied here, which in essence reduces the proof to a simple application

of Minkowski's inequality. For item (a), note that lemma B.15 gives the following result

$$\begin{aligned}
\mathrm{E}\big[\big|Z^n_{hq:t}\big|^\nu\big]^{1/\nu} &= \mathrm{E}\big[\big|X^n_{hq:t} - \mathrm{E}\big[X^n_{hq:t}\big]\big|^\nu\big]^{1/\nu} && \text{(B.39a)}\\
&\leq \mathrm{E}\big[\big|X^n_{hq:t}\big|^\nu\big]^{1/\nu} + \mathrm{E}\big[\big|\mathrm{E}\big[X^n_{hq:t}\big]\big|^\nu\big]^{1/\nu} && \text{(B.39b)}\\
&= O\big(|b_1 b_2|^{(2-\nu)/2\nu}\big) + O\Big(\sqrt{b_1 b_2}\Big) && \text{(B.39c)}\\
&= O\big(|b_1 b_2|^{(2-\nu)/2\nu}\big)\,. && \text{(B.39d)}
\end{aligned}$$

Item (b) now follows from item (a) and lemma C.8, due to the following inequality,

$$\begin{aligned}
\mathrm{E}[|Z^n_{\bar{m}:t}(\boldsymbol{a})|^\nu]^{1/\nu} &= \mathrm{E}\Big[\Big|\textstyle\sum_{h=1}^m \sum_{q=1}^5 a_{hq} Z^n_{hq:t}\Big|^\nu\Big]^{1/\nu} && \text{(B.40a)}\\
&\leq \sum_{h=1}^m \sum_{q=1}^5 \big|a_{hq}\big|\,\mathrm{E}\big[\big|Z^n_{hq:t}\big|^\nu\big]^{1/\nu} && \text{(B.40b)}\\
&\leq \sum_{h=1}^m \sum_{q=1}^5 A_{\bar{m}} \cdot O\big(|b_1 b_2|^{(2-\nu)/2\nu}\big) && \text{(B.40c)}\\
&= O\big(m\,|b_1 b_2|^{(2-\nu)/2\nu}\big)\,. && \text{(B.40d)}
\end{aligned}$$

where $A_{\bar{m}}$ is the maximum of $\big|a_{hq}\big|$. □

### B.2.3 The asymptotic results — final part

This section will present the final steps toward the verification of the fourth requirement of the Klimko-Nelson approach for the case where $m \to \infty$ and $\boldsymbol{b} \to \mathbf{0}^+$ when $n \to \infty$. Note that theorem B.20 (the main theorem) requires both a large block - small block argument and a truncation argument, and the technical details related to these components will be taken care of in lemma B.18 and corollary B.19.

The large block - small block argument requires that quite a few components must be verified to be asymptotically negligible. The following lemma, which extends an argument encountered in the proof of Masry and Tjøstheim [1995, Lemma 4.3(b)], shows that the asymptotic negligibility of all the 'off the diagonal' components can be taken care of in one operation.

**Lemma B.18.** *When $Y_t$ satisfies assumption 2.1, when $n$, $m$ and $\boldsymbol{b}$ are as specified in assumption 2.3, and when $u_{hq:\boldsymbol{b}}(\boldsymbol{w})$ and $K(\boldsymbol{w})$ are as given in definitions B.7 and B.9 —*

*then the random variables* $Z^n_{\bar{m}:t}(\boldsymbol{a})$ *from definition B.13 satisfies*

$$\frac{1}{n}\sum_{\substack{i,k=1\\ i\neq k}}^{n} \left|\mathrm{E}[Z^n_{\bar{m}:i}(\boldsymbol{a})\cdot Z^n_{\bar{m}:k}(\boldsymbol{a})]\right| = o(1)\,. \tag{B.41}$$

*Proof.* Assumption 2.1(a), i.e. the strict stationarity of $\{Y_t\}_{t\in\mathbb{Z}}$, implies that the double sum in eq. (B.41) can be reduced to a single sum, i.e.

$$\frac{1}{n}\sum_{\substack{i,k=1\\ i\neq k}}^{n} \left|\mathrm{E}[Z^n_{\bar{m}:i}(\boldsymbol{a})\cdot Z^n_{\bar{m}:k}(\boldsymbol{a})]\right| = 2\sum_{\ell=1}^{n-1}\left(1-\frac{\ell}{n}\right) I^n_{\bar{m}:\ell}(\boldsymbol{a}), \tag{B.42}$$

where the terms $I^n_{\bar{m}:\ell}(\boldsymbol{a})$ are given by

$$\begin{aligned}
I^n_{\bar{m}:\ell}(\boldsymbol{a}) &\coloneqq \left|\mathrm{E}[Z^n_{\bar{m}:0}(\boldsymbol{a})\cdot Z^n_{\bar{m}:\ell}(\boldsymbol{a})]\right| && \text{(B.43a)}\\
&= \left|\mathrm{E}\left[\sum_{h=1}^{m}\sum_{q=1}^{5} a_{hq} Z^n_{hq:0}\cdot \sum_{j=1}^{m}\sum_{r=1}^{5} a_{jr} Z^n_{jr:\ell}\right]\right| && \text{(B.43b)}\\
&= \left|\sum_{h=1}^{m}\sum_{j=1}^{m}\sum_{q=1}^{5}\sum_{r=1}^{5} a_{hq}a_{jr}\,\mathrm{E}\left[Z^n_{hq:0}\cdot Z^n_{jr:\ell}\right]\right| && \text{(B.43c)}\\
&\leq \sum_{h=1}^{m}\sum_{j=1}^{m}\sum_{q=1}^{5}\sum_{r=1}^{5} |a_{hq}||a_{jr}|\, I^n_{hqjr:\ell}, && \text{(B.43d)}
\end{aligned}$$

where $I^n_{hqjr:\ell} \coloneqq \left|\mathrm{E}\left[Z^n_{hq:0}\cdot Z^n_{jr:\ell}\right]\right| = \left|\mathrm{Cov}\left(X^n_{hq:0}, X^n_{jr:\ell}\right)\right|$.

Introducing integers $k_n$ (to be specified later on) such that $k_n\to\infty$ and $k_n m^2 b_1 b_2 \to 0$ as $n\to\infty$, eq. (B.42) can be written as the sum of the following three sums,

$$J_1 \coloneqq 2\sum_{\ell=1}^{m}(1-\ell/n)\, I^n_{\bar{m}:\ell}(\boldsymbol{a}), \tag{B.44a}$$

$$J_2 \coloneqq 2\sum_{\ell=m+1}^{k_n+m}(1-\ell/n)\, I^n_{\bar{m}:\ell}(\boldsymbol{a}), \tag{B.44b}$$

$$J_3 \coloneqq 2\sum_{\ell=k_n+m+1}^{n-1}(1-\ell/n)\, I^n_{\bar{m}:\ell}(\boldsymbol{a}). \tag{B.44c}$$

From the definition of $I^n_{\overline{m}:\ell}(\boldsymbol{a})$ it is seen that in $J_1$ there will be some overlap between those $Y_t$ that are a part of $Z^n_{\overline{m}:0}(\boldsymbol{a})$ and those that are a part of $Z^n_{\overline{m}:\ell}(\boldsymbol{a})$, and moreover that this will not be the case for the two sums $J_2$ and $J_3$.

Equations (B.43d) and (B.44a) implies that a squeeze argument can be used when dealing with $J_1$, i.e.

$$0 \leq J_1 \leq 2 \cdot \left( \max_{\substack{h\in\{1,\ldots,m\}\\ q\in\{1,\ldots,5\}}} \left|a_{hq}\right|^2 \right) \cdot \sum_{\ell=1}^{m} \sum_{h=1}^{m} \sum_{j=1}^{m} \sum_{q=1}^{5} \sum_{r=1}^{5} \left|\mathrm{Cov}\left(X^n_{hq:0}, X^n_{jr:\ell}\right)\right|, \tag{B.45}$$

and corollary B.16 can be used to determine how the summand behaves in the limit. Table 1, page 19, shows that the bivariate case never occurs, that $h$ must be equal to $\ell$ or $j+\ell$ in order for a trivariate case to occur, and that the rest of the cases must be tetravariate. It is not hard (but a bit tedious) to explicitly compute the number of trivariate terms that occur in eq. (B.45), but for the present asymptotic analysis it is sufficient to note that the number of trivariate terms is of order $m^2$, whereas the number of tetravariate terms is of order $m^3$. Corollary B.16 thus gives that the bivariate and tetravariate parts of the bound for $J_1$ respectively are $O(m^2(b_1 \wedge b_2))$ and $O(m^3 b_1 b_2)$.

$J_1 = o(1)$ now follows from assumption 2.3(c) and the following two simple observations;

$$m^2(b_1 \wedge b_2) \leq m^2(b_1 \vee b_2)\,, \tag{B.46a}$$

$$m^3 b_1 b_2 \leq m^{-1} \cdot m^4 (b_1 \vee b_2)^2 = m^{-1} \cdot \left(m^2\,(b_1 \vee b_2)\right)^2. \tag{B.46b}$$

For $J_2$, a squeeze similar to the one in eq. (B.45) can be used. The situation becomes simpler since $\ell > M$ ensures that only the tetravariate case is present, and the order of $J_2$ becomes

$$J_2 = O(k_n m^2 b_1 b_2)\,. \tag{B.47}$$

Since $k_n m^2 b_1 b_2 \to 0$ (with a choice of $k_n$ to be specified below), it follows that $J_2 = o(1)$.

For $J_3$, the Corollary of Lemma 2.1 in Davydov [1968] will be used to get an upper bound on $I^n_{\overline{m}:\ell}(\boldsymbol{a})$, such that a squeeze-argument can be used for $J_3$ too. The requirements needed for Davydov's result are covered as follows: The strong mixing requirement is covered by assumption 2.1, and (for a given $m$ and $\boldsymbol{b}$) the requirement about finite expectations follows from corollary B.17(b).

The $\sigma$-algebras to be used follows from the comment stated after eq. (C.33), i.e. that $Z^n_{\overline{m}:0}(\boldsymbol{a}) \in \mathcal{F}^m_0$, whereas $Z^n_{\overline{m}:\ell}(\boldsymbol{a}) \in \mathcal{F}^{\ell+m}_{\ell} \subset \mathcal{F}^{\infty}_{m+(\ell-m)}$. Thus, for $\ell > k_n + m$, the following

bound is obtained on $I^n_{\bar{m}:\ell}(\boldsymbol{a})$,

$$
\begin{aligned}
I^n_{\bar{m}:\ell}(\boldsymbol{a}) &= \left|\mathrm{E}[Z^n_{\bar{m}:0}(\boldsymbol{a}) \cdot Z^n_{\bar{m}:\ell}(\boldsymbol{a})]\right| && \text{(B.48a)}\\
&= \left|\mathrm{E}[Z^n_{\bar{m}:0}(\boldsymbol{a}) \cdot Z^n_{\bar{m}:\ell}(\boldsymbol{a})] - \mathrm{E}[Z^n_{\bar{m}:0}(\boldsymbol{a})] \cdot \mathrm{E}[Z^n_{\bar{m}:\ell}(\boldsymbol{a})]\right| && \text{(B.48b)}\\
&\leq 12 \left(\mathrm{E}[|Z^n_{\bar{m}:0}(\boldsymbol{a})|^\nu]\right)^{1/\nu} \cdot \left(\mathrm{E}[|Z^n_{\bar{m}:\ell}(\boldsymbol{a})|^\nu]\right)^{1/\nu} \cdot [\alpha(\ell-m)]^{1-1/\nu-1/\nu} && \text{(B.48c)}\\
&= 12 \left(\left(\mathrm{E}[|Z^n_{\bar{m}:0}(\boldsymbol{a})|^\nu]\right)^{1/\nu}\right)^2 \cdot [\alpha(\ell-m)]^{1-2/\nu} && \text{(B.48d)}\\
&= 12 \left(O\big(m\,|b_1 b_2|^{(2-\nu)/2\nu}\big)\right)^2 \cdot [\alpha(\ell-m)]^{1-2/\nu} && \text{(B.48e)}\\
&\leq \mathcal{C} \cdot m^2 \cdot |b_1 b_2|^{(2-\nu)/\nu} \cdot [\alpha(\ell-m)]^{1-2/\nu}\,, && \text{(B.48f)}
\end{aligned}
$$

where eq. (B.48b) follows since the mean of $Z^n_{\bar{m}:t}(\boldsymbol{a})$ by construction is zero, where eq. (B.48c) is Davydov's result, where eq. (B.48d) use the strict stationarity of the process $\{Y_t\}$, where eq. (B.48e) is due to corollary B.17(b), and finally eq. (B.48f) is an equivalent statement, using a suitable constant $\mathcal{C}$ to express the upper bound.

A squeeze for $J_3$ can now be stated in the following manner

$$
0 \leq J_3 \leq \mathcal{C}_3 \cdot \sum_{j=k_n+1}^{\infty} \left(m^2 \cdot |b_1 b_2|^{(2-\nu)/\nu}\right) \cdot [\alpha(j)]^{1-2/\nu}\,, \tag{B.49}
$$

where $\mathcal{C}_3$ is a constant, where the index has been shifted by introducing $j = \ell - m$, and where the sum from eq. (B.44c) has been extended to infinity (adding only non-negative summands).

A comparison of eq. (B.49) with the finiteness requirement that the strong mixing coefficients should satisfy, see assumption 2.1(b), indicates that if $j^a \geq m^2 \cdot |b_1 b_2|^{(2-\nu)/\nu}$ for $j \geq k_n + 1$, then that could be used to get a new upper bound in eq. (B.49). Taking the $a^{\text{th}}$ root on both sides, it is clear that the desired inequality can be obtained when $k_n + 1 = \left\lceil m^{2/a} \cdot |b_1 b_2|^{(2-\nu)/a\nu} \right\rceil$, which gives the new bound

$$
0 \leq J_3 \leq \mathcal{C}_3 \cdot \sum_{j=k_n+1}^{\infty} j^a \,[\alpha(j)]^{1-2/\nu}\,, \tag{B.50}
$$

and if $k_n \to \infty$ when $n \to \infty$, the finiteness assumption from assumption 2.1(b) gives that $J_3 = o(1)$.

Finally, lemma C.4 verifies that $k_n$ satisfies the two limits $k_n m^2 b_1 b_2 \to 0$ (needed for the $J_2$-term) and $k_n \to \infty$ (needed for the $J_3$-term). Altogether, this shows that eq. (B.41) can be rewritten as $J_1 + J_2 + J_3$, all of which are $o(1)$, and the proof is complete. □

The following observations are needed in the truncation argument of theorem B.20.

**Corollary B.19.** *When $Y_t$ satisfies assumption 2.1, when $n$, $m$ and $\boldsymbol{b}$ are as specified in assumption 2.3, and with $W_{\bar{m}:\boldsymbol{b}} = \bigoplus_{h=1}^{m} W_{h:\boldsymbol{b}}$ and $\boldsymbol{a} = \boldsymbol{a}_{\bar{m}} = [\boldsymbol{a}_1, \dots, \boldsymbol{a}_m]'$ (with $\boldsymbol{a}_h \in \mathbb{R}^5$) as given in definition B.10, then the random variable $Z^n_{\bar{m}:t}(\boldsymbol{a})$ from definition B.13 satisfies*

*(a)* $\operatorname{Var}(Z^n_{\bar{m}:t}(\boldsymbol{a})) = \boldsymbol{a}'_{\bar{m}} W_{\bar{m}:\boldsymbol{b}} \boldsymbol{a}_{\bar{m}} + O(m^2 \cdot (b_1 \vee b_2)) = \sum_{h=1}^{m} \boldsymbol{a}'_h W_{h:\boldsymbol{b}} \boldsymbol{a}_h + O(m^2 \cdot (b_1 \vee b_2)) = O(m)$.

*Furthermore, with $r := r_n$ a sequence of integers that goes to $\infty$ when $n \to \infty$, and for a given threshold value $L$, the following holds for the random variables $\eta_{1:r} := \sum_{t=1}^{r} Z^n_{\bar{m}:t}(\boldsymbol{a})$, $\eta^{\leq L}_{1:r} := \sum_{t=1}^{r} Z^{n|\leq L}_{\bar{m}:t}(\boldsymbol{a})$ and $\eta^{>L}_{1:r} := \sum_{t=1}^{r} Z^{n|>L}_{\bar{m}:t}(\boldsymbol{a})$.*

*(b)* $\operatorname{Var}(\eta_{1:r}) = r \cdot \{\sum_{h=1}^{m} \boldsymbol{a}'_h W_{h:\boldsymbol{b}} \boldsymbol{a}_h + o(1)\}$.

*(c) When $L$ is large enough,* $\operatorname{Var}(\eta^{\leq L}_{1:r}) = r \cdot \{\sum_{h=1}^{m} \boldsymbol{a}'_h W_{h:\boldsymbol{b}} \boldsymbol{a}_h + o(1)\}$ *and* $\operatorname{Var}(\eta^{>L}_{1:r}) = r \cdot o(1)$.

*Proof.* For item (a), note that it follows from definitions B.12 and B.13 that

$$\operatorname{Var}(Z^n_{\bar{m}:t}(\boldsymbol{a})) = \sum_{h=1}^{m}\sum_{j=1}^{m}\sum_{q=1}^{5}\sum_{r=1}^{5} a_{hq} a_{jr} \operatorname{Cov}\big(X^n_{hq:t}, X^n_{jr:t}\big) \tag{B.51a}$$

$$= \sum_{h=1}^{m}\sum_{q=1}^{5}\sum_{r=1}^{5} a_{hq} a_{hr} \operatorname{Cov}\big(X^n_{hq:t}, X^n_{hr:t}\big) + \sum_{\substack{h,j=1 \\ h\neq j}}^{m}\sum_{q=1}^{5}\sum_{r=1}^{5} a_{hq} a_{jr} \operatorname{Cov}\big(X^n_{hq:t}, X^n_{jr:t}\big). \tag{B.51b}$$

The bivariate case of corollary B.16 can be applied to the 'diagonal part' of the sum in eq. (B.51b), whereas the trivariate and tetravariate cases can be applied to the 'off-diagonal part'. The 'diagonal part' can thus be written as the sum of $\sum_{h=1}^{m}\sum_{q=1}^{5}\sum_{r=1}^{5} a_{hq} a_{hr} u_{hq:\boldsymbol{b}}(\boldsymbol{v})\, u_{hr:\boldsymbol{b}}(\boldsymbol{v})\, g_h(\boldsymbol{v}) \int_{\mathbb{R}^2} K(\boldsymbol{w})^2 \,\mathrm{d}\boldsymbol{w}$ (which is equal to $\boldsymbol{a}' W_{\bar{m}:\boldsymbol{b}} \boldsymbol{a} = \sum_{h=1}^{m} \boldsymbol{a}'_h W_{h:\boldsymbol{b}} \boldsymbol{a}_h$) and a sum that is $O(m \cdot (b_1 \vee b_2))$. For the 'off-diagonal part' the result is $O(m^2 \cdot (b_1 \wedge b_2))$. Both of these asymptotically negligible terms are covered by $O(m^2 \cdot (b_1 \vee b_2))$, and this gives the two first equalities of item (a). The last equality follows since the summands $\boldsymbol{a}'_h W_{h:\boldsymbol{b}} \boldsymbol{a}_h$ are finite.

For item (b), note that the variance can be expressed as

$$\operatorname{Var}(\eta_{1:r}) = \sum_{i=1}^{r} \operatorname{Var}(Z^n_{\bar{m}:i}(\boldsymbol{a})) + \sum_{\substack{i,k=1 \\ i\neq k}}^{r} \mathrm{E}[Z^n_{\bar{m}:i}(\boldsymbol{a}) \cdot Z^n_{\bar{m}:k}(\boldsymbol{a})]\,. \tag{B.52}$$

The 'on diagonal' part of this sum equals $r \cdot \operatorname{Var}(Z^n_{\bar{m}:1}(\boldsymbol{a}))$ due to assumption 2.1(a), while the 'off diagonal' part due to lemma B.18 becomes $r \cdot o(1)$. Together with the result from item (a), this gives the statement in item (b).

The truncated cases in item (c) use the same arguments as those encountered in item (b), with the effect that the $u_{hq:\boldsymbol{b}}(\boldsymbol{v})\, u_{hr:\boldsymbol{b}}(\boldsymbol{v})$ that occurs in $W_{h:\boldsymbol{b}}$ either are replaced by $u_{hq:\boldsymbol{b}}(\boldsymbol{v})^{\leq L}\, u_{hr:\boldsymbol{b}}(\boldsymbol{v})^{\leq L}$ or by $u_{hq:\boldsymbol{b}}(\boldsymbol{v})^{>L}\, u_{hr:\boldsymbol{b}}(\boldsymbol{v})^{>L}$. Lemma B.8(b) gives that $u_{hq:\boldsymbol{b}}(\boldsymbol{v})^{\leq L} = u_{hq:\boldsymbol{b}}(\boldsymbol{v})$ when $L$ is large enough (and thus $u_{hq:\boldsymbol{b}}(\boldsymbol{v})^{>L} = 0$), which completes the proof. □

The main theorem can now be stated, i.e. this result can be used to verify the fourth requirement of the Klimko-Nelson approach for the penalty function $Q_{\overline{m}:n}(\boldsymbol{\theta}_{\overline{m}:\boldsymbol{b}})$, from which it follows an asymptotic normality result for $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}}$, that finally gives the asymptotic normality result of $\widehat{f}^{m}_{\boldsymbol{v}}(\omega)$. (Confer appendix B.2.3 for an interpretation of the $m$ that occurs in the limiting distributions.)

**Theorem B.20.** *For a given point $\boldsymbol{v} = (v_1, v_2)$: When $Y_t$ satisfies assumptions 2.1 and 2.2, when $n$, $m$ and $\boldsymbol{b}$ are as specified in assumption 2.3, and with $W_{\overline{m}:\boldsymbol{b}} = \bigoplus_{h=1}^{m} W_{h:\boldsymbol{b}}$ and $\boldsymbol{a} = \boldsymbol{a}_{\overline{m}} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m]'$ (with $\boldsymbol{a}_h \in \mathbb{R}^5$) as given in definition B.10, then the random variables $\mathcal{Y}^{n}_{\overline{m}}(\boldsymbol{a})$ and $\boldsymbol{\mathcal{Y}}^{n}_{\overline{m}}$ from definition B.13 will for small $\boldsymbol{b}$ and large $m$ and $n$ satisfy*

*(a) $n^{-1/2}\, \mathcal{Y}^{n}_{\overline{m}}(\boldsymbol{a}) \xrightarrow{d} \mathrm{N}(0, \sum_{h=1}^{m} \boldsymbol{a}_h' W_{h:\boldsymbol{b}} \boldsymbol{a}_h)$, i.e. asymptotically univariate normal.*

*(b) $n^{-1/2}\, \boldsymbol{\mathcal{Y}}^{n}_{\overline{m}} \xrightarrow{d} \mathrm{N}(\boldsymbol{0}, \bigoplus_{h=1}^{m} W_{h:\boldsymbol{b}})$, i.e. asymptotically $5m$-variate normal.*

*Proof.* For the proof of item (a), note the following connection between $\mathcal{Y}^{n}_{\overline{m}}(\boldsymbol{a})$ and $Z^{n}_{\overline{m}:t}(\boldsymbol{a})$ which follows directly from definitions B.12 and B.13,

$$\mathcal{Y}^{n}_{\overline{m}}(\boldsymbol{a}) = \sum_{h=1}^{m}\sum_{q=1}^{5} a_{hq}\, \mathcal{Y}^{n}_{hq} = \sum_{h=1}^{m}\sum_{q=1}^{5} a_{hq} \left[\sum_{t=1}^{n} Z^{n}_{hq:t}\right] = \sum_{t=1}^{n} \left[\sum_{h=1}^{m}\sum_{q=1}^{5} a_{hq} Z^{n}_{hq:t}\right] = \sum_{t=1}^{n} Z^{n}_{\overline{m}:t}(\boldsymbol{a}). \tag{B.53a}$$

A large block - small block argument can be used to analyse this, i.e. the index set $\{1, \ldots, n\}$ will be partitioned into large blocks and small blocks, such that $\mathcal{Y}^{n}_{\overline{m}}(\boldsymbol{a})$ can be expressed as the sum of $S^{(1)}_n$, $S^{(2)}_n$ and $S^{(3)}_n$ (to be defined below). The asymptotic distribution of $\mathcal{Y}^{n}_{\overline{m}}(\boldsymbol{a})$ will be shown to coincide with the asymptotic distribution of $S^{(1)}_n$, the summands of $S^{(1)}_n$ will be shown to be asymptotically independent, and finally the Lindeberg conditions for asymptotic normality of $S^{(1)}_n$ will be verified.

Use $\ell$, $r$, and $s$ from lemma C.3(c) to divide the indexing set $\{1, \ldots, n\}$ into $2\ell + 1$ subsets of large blocks and small blocks (and one reminder block), defined as follows

$$\mathcal{A}_j \coloneqq \{(j-1)(r+s)+1, \ldots, (j-1)(r+s)+r\}, \text{ for } j = 1, \ldots, \ell, \tag{B.54a}$$

$$\mathcal{B}_j \coloneqq \{(j-1)(r+s)+r+1, \ldots, j(r+s)\}, \text{ for } j = 1, \ldots, \ell, \tag{B.54b}$$

$$\mathcal{C}_\ell \coloneqq \begin{cases} \{\ell(r+s)+1, \ldots, n\} & \text{when } \ell(r+s) < n, \\ \emptyset & \text{when } \ell(r+s) = n. \end{cases} \tag{B.54c}$$

In order to avoid iterated sums later on, introduce the following unions,

$$\mathcal{A}^\circ \coloneqq \bigcup_{j=1}^{\ell} \mathcal{A}_j, \qquad \mathcal{B}^\circ \coloneqq \bigcup_{j=1}^{\ell} \mathcal{B}_j. \tag{B.55a}$$

Note that the number of elements in $\mathcal{A}^\circ$ and $\mathcal{B}^\circ$ will be $\ell r$ and $\ell s$ respectively. The number of elements in $\mathcal{C}_\ell$ will be $n - \ell(r+s)$, and this can vary between 0 and $r + s - 1 < 2r$.

Use these subsets of $\{1, \ldots, n\}$ to define the following variables,

$$\eta_j \coloneqq \sum_{t \in \mathcal{A}_j} Z^n_{\overline{m}:t}(\boldsymbol{a}), \text{ for } j = 1, \ldots, \ell, \qquad S^{(1)}_n \coloneqq \sum_{j=1}^{\ell} \eta_j = \sum_{t \in \mathcal{A}^\circ} Z^n_{\overline{m}:t}(\boldsymbol{a}), \tag{B.56a}$$

$$\xi_j \coloneqq \sum_{t \in \mathcal{B}_j} Z^n_{\overline{m}:t}(\boldsymbol{a}), \text{ for } j = 1, \ldots, \ell, \qquad S^{(2)}_n \coloneqq \sum_{j=1}^{\ell} \xi_j = \sum_{t \in \mathcal{B}^\circ} Z^n_{\overline{m}:t}(\boldsymbol{a}), \tag{B.56b}$$

$$\zeta_\ell \coloneqq \sum_{t \in \mathcal{C}_\ell} Z^n_{\overline{m}:t}(\boldsymbol{a}), \qquad S^{(3)}_n \coloneqq \zeta_\ell, \tag{B.56c}$$

such that

$$n^{-1/2} \, \text{Ϙ}^n_{\overline{m}}(\boldsymbol{a}) = n^{-1/2} \left\{ S^{(1)}_n + S^{(2)}_n + S^{(3)}_n \right\}. \tag{B.57}$$

The expectation of these quantities are by construction equal to zero, which gives

$$\operatorname{Var}(n^{-1/2} \, \text{Ϙ}^n_{\overline{m}}(\boldsymbol{a})) = \frac{1}{n} \operatorname{E}[\text{Ϙ}^n_{\overline{m}}(\boldsymbol{a}) \cdot \text{Ϙ}^n_{\overline{m}}(\boldsymbol{a})] = \frac{1}{n} \sum_{p=1}^{3} \sum_{q=1}^{3} \operatorname{E}[S^{(p)}_n \cdot S^{(q)}_n]. \tag{B.58}$$

When $p \neq q$, there will be no overlap between the indexing sets that occur in the two sums, and the following inequality, here illustrated by the case $p = 1$ and $q = 2$, is obtained

$$\left| \frac{1}{n} \operatorname{E}[S^{(1)}_n \cdot S^{(2)}_n] \right| = \left| \frac{1}{n} \operatorname{E}\left[ \left( \sum_{i \in \mathcal{A}^\circ} Z^n_{\overline{m}:i}(\boldsymbol{a}) \right) \cdot \left( \sum_{k \in \mathcal{B}^\circ} Z^n_{\overline{m}:k}(\boldsymbol{a}) \right) \right] \right| \tag{B.59a}$$

$$\leq \frac{1}{n} \sum_{i \in \mathcal{A}^\circ} \sum_{k \in \mathcal{B}^\circ} \left| \operatorname{E}[Z^n_{\overline{m}:i}(\boldsymbol{a}) \cdot Z^n_{\overline{m}:k}(\boldsymbol{a})] \right| \tag{B.59b}$$

$$\leq \frac{1}{n} \sum_{\substack{i,k=1 \\ i \neq k}}^{n} \left| \operatorname{E}[Z^n_{\overline{m}:i}(\boldsymbol{a}) \cdot Z^n_{\overline{m}:k}(\boldsymbol{a})] \right|. \tag{B.59c}$$

Lemma B.18 thus gives that the expectation of all the cross-terms are asymptotically negligible.

For the case $p = q = 2$, i.e. the small blocks, the same strategy as in eq. (B.59) shows that the internal cross-terms are asymptotically negligible. Corollary B.19(a) states that the remaining summands all are $O(m)$, which results in the following bound

$$\frac{1}{n}\,\mathrm{E}[S_n^{(2)} \cdot S_n^{(2)}] = \frac{1}{n}\sum_{i,k\in\mathcal{B}^\circ} \mathrm{E}[Z^n_{\bar{m}:i}(\boldsymbol{a}) \cdot Z^n_{\bar{m}:k}(\boldsymbol{a})] \tag{B.60a}$$

$$= \frac{1}{n}\sum_{i\in\mathcal{B}^\circ} \mathrm{E}[Z^n_{\bar{m}:i}(\boldsymbol{a}) \cdot Z^n_{\bar{m}:i}(\boldsymbol{a})] + \frac{1}{n}\sum_{\substack{i,k\in\mathcal{B}^\circ \\ i\neq k}} \mathrm{E}[Z^n_{\bar{m}:i}(\boldsymbol{a}) \cdot Z^n_{\bar{m}:k}(\boldsymbol{a})] \tag{B.60b}$$

$$= \frac{1}{n}\sum_{i\in\mathcal{B}^\circ} O(m) + o(1) \tag{B.60c}$$

$$= O\Big(\frac{m\ell s}{n}\Big)\,. \tag{B.60d}$$

For the case $p = q = 3$, i.e. the residual block, a similar argument gives

$$\frac{1}{n}\,\mathrm{E}[S_n^{(3)} \cdot S_n^{(3)}] = O\Big(\frac{m\,(n-\ell(r+s))}{n}\Big) < O\Big(\frac{mr}{n}\Big)\,. \tag{B.61}$$

Lemma C.3(c) ensures that $(m\ell s)/n$ and $mr/n$ goes to zero, so the terms investigated in eq. (B.60) and eq. (B.61) are asymptotically negligible. This implies that $n^{-1/2}(\Qoppa^n_{\bar{m}}(\boldsymbol{a}) - S_n^{(1)}) \Rightarrow 0$, and Billingsley [2012, Theorem 25.4] states that there thus is a common limiting distribution for $n^{-1/2}\,\Qoppa^n_{\bar{m}}(\boldsymbol{a})$ and $n^{-1/2}\,S_n^{(1)}$.

The arguments used for $S_n^{(2)}$ also gives the simple observation below, which is needed later on,

$$\mathrm{Var}(n^{-1/2}\,S_n^{(1)}) = \frac{1}{n}\sum_{j=1}^{\ell} \mathrm{Var}\big(\eta_j\big) + o(1)\,. \tag{B.62}$$

The next step is to show that the random variables $\eta_j$ are asymptotically independent, which formulated relative to the characteristic functions corresponds to showing

$$\left|\mathrm{E}[\exp(itS_n^{(1)})] - \prod_{j=1}^{\ell} \mathrm{E}\big[\exp\big(it\eta_j\big)\big]\right| \to 0. \tag{B.63}$$

The validity of this statement follows from Lemma 1.1 in Volkonskii and Rozanov [1959, p. 180], by introducing random variables $V_j = \exp\big(it\eta_j\big)$, for $j = 1, \ldots, \ell$. By construction, the $V_j$ trivially satisfies the requirement $\big|V_j\big| \leq 1$, so it only remains to identify the corresponding $\sigma$-algebras and the distance between them. From the definitions of $\eta_j$, $\mathcal{A}_j$ and $Z^n_{\bar{m}:t}(\boldsymbol{a})$, it is easy to see that $V_j \in \mathcal{F}^{(j-1)(r+s)+r+m}_{(j-1)(r+s)+1}$, and from this it follows that the distance between the highest index in the $\sigma$-algebra corresponding to $V_j$ and the lowest index in the $\sigma$-algebra corresponding to $V_{j+1}$, is given by

$$\vartheta \coloneqq \{((j+1)-1)(r+s)+1\} - \{(j-1)(r+s)+r+m\} = s-m+1. \tag{B.64}$$

Assumption 2.3(f), i.e. $m = o(s)$, ensures that there (asymptotically) will be no overlap between these $\sigma$-algebras, and the result from Volkonskii and Rozanov [1959] thus gives $16(\ell-1)\alpha(\vartheta)$ as an upper bound on the left side of eq. (B.63). Lemma C.3(c) says that this bound goes to zero, which shows that the $\eta_j$ are asymptotically independent.

It remains to verify the Lindeberg condition, for which an expression for $\mathfrak{s}^2_\ell \coloneqq \sum_{j=1}^{\ell} \operatorname{Var}\big(\eta_j\big)$ is needed. From assumption 2.1(a) and corollary B.19(b), it follows that

$$\mathfrak{s}^2_\ell = \sum_{j=1}^{\ell} \operatorname{Var}\big(\eta_j\big) = \ell \cdot \operatorname{Var}(\eta_1) = \ell \cdot r \cdot \left\{ \sum_{h=1}^{m} \boldsymbol{a}_h' W_{h:\boldsymbol{b}} \boldsymbol{a}_h + o(1) \right\}, \tag{B.65}$$

and assuming $\mathfrak{s}^2_\ell > 0$, the condition to verify is

$$\forall\ \epsilon > 0 \qquad \lim_{n\to\infty} \sum_{j=1}^{\ell} \frac{1}{\mathfrak{s}^2_\ell} \operatorname{E}\big[\eta_j^2 \cdot \mathbb{1}\big\{\big|\eta_j\big| \geq \epsilon\sqrt{\mathfrak{s}^2_\ell}\big\}\big] \longrightarrow 0. \tag{B.66}$$

This holds trivially if the sets occurring in the indicator functions, i.e. $\big\{\big|\eta_j\big| \geq \epsilon\sqrt{\mathfrak{s}^2_\ell}\big\}$, becomes empty when $n$ is large enough. It is thus of interest to see if an upper bound for $\big|\eta_j\big|$ can be found, and if the limit of this upper bound becomes smaller than the limit of the right-hand side $\epsilon\sqrt{\mathfrak{s}^2_\ell}$.

Keeping in mind the definitions of $X^n_{hq:t}$, $Z^n_{hq:t}$ and $\eta_j$, see eqs. (B.30a), (B.32a) and (B.56a), it is clear that an upper bound for $\big|\eta_j\big|$ might be deduced from,

$$\big|\eta_j\big| = \left| \sum_{t\in\mathcal{A}_j} \sum_{h=1}^{m} \sum_{q=1}^{5} a_{hq} Z^n_{hq:t} \right| \leq \sum_{t\in\mathcal{A}_j} \sum_{h=1}^{m} \sum_{q=1}^{5} \big|a_{hq}\big|\,\big|Z^n_{hq:t}\big|, \tag{B.67a}$$

$$\big|Z^n_{hq:t}\big| = \big|X^n_{hq:t} - \operatorname{E}\big[X^n_{hq:t}\big]\big| \leq \big|X^n_{hq:t}\big| + O\Big(\sqrt{b_1 b_2}\Big), \tag{B.67b}$$

$$\big|X^n_{hq:t}\big| = \left| \sqrt{b_1 b_2} \cdot \frac{1}{b_1 b_2} K_h\left( \frac{Y_{t+h} - v_1}{b_1}, \frac{Y_t - v_2}{b_2} \right) u_{h:\boldsymbol{b}}(\boldsymbol{Y}_{h:t}) \right|. \tag{B.67c}$$

If all of the functions $u_{hq:\boldsymbol{b}}(\boldsymbol{w})$ are bounded, or if the kernel functions $K_{h:\boldsymbol{b}}(\boldsymbol{w}-\boldsymbol{v})$ have bounded support, then the present framework will be sufficient to reach the desired conclusion. However, no such conditions are assumed, and a truncation argument must thus be introduced in order to deal with this problem — in particular, the expression $\wp_{\bar{m}}^{n}(\boldsymbol{a}) = \wp_{\bar{m}}^{n|\leq L}(\boldsymbol{a}) + \wp_{\bar{m}}^{n|>L}(\boldsymbol{a})$ will be used.

Lemma B.8(a) implies that a large enough value for the threshold $L$ will ensure that all constructions and arguments based upon the ordinary functions $u_{hq:\boldsymbol{b}}(\boldsymbol{w})$ also works nicely for the truncated functions $u_{hq:\boldsymbol{b}}(\boldsymbol{w})^{\leq L}$ and $u_{hq:\boldsymbol{b}}(\boldsymbol{w})^{>L}$. With regard to the limiting distributions, first note that $n^{-1/2}\,\wp_{\bar{m}}^{n|>L}(\boldsymbol{a})$ and $n^{-1/2}\,S_n^{(1)|>L}$ shares the same limiting distribution, and then observe that the upper truncated versions of eqs. (B.62) and (B.65) together with the result from corollary B.19(c), gives the following bound when $L$ is large enough:

$$\operatorname{Var}\left(n^{-1/2}\,S_n^{(1)|>L}\right) = \frac{1}{n}\sum_{j=1}^{\ell}\operatorname{Var}\left(\eta_j^{>L}\right) + o(1) = \frac{\ell r}{n}\cdot o(1)\,. \tag{B.68}$$

Since $\ell r \asymp n$, it follows that $n^{-1/2}\,\wp_{\bar{m}}^{n|>L}(\boldsymbol{a}) \Rightarrow 0$, so the limiting distributions of $n^{-1/2}\,\wp_{\bar{m}}^{n}(\boldsymbol{a})$ and $n^{-1/2}\,\wp_{\bar{m}}^{n|\leq L}(\boldsymbol{a})$ coincide when $L$ is large enough.[9] Next, observe that the random variable $\left|\eta_j^{\leq L}\right|$ obviously will have an upper bound, since the truncated polynomial $u_{hq:\boldsymbol{b}}(\boldsymbol{w})^{\leq L}$ will occur in the lower truncated version of eq. (B.67). Since the kernel function $K(\boldsymbol{w})$ by definition is bounded by some constant $\mathcal{K}$, it follows that $\left|\eta_j^{\leq L}\right|$ is bounded by

$$\left|\eta_j^{\leq L}\right| \leq 5rm\left(\max\left|a_{hq}\right|\right)\left(\frac{\mathcal{K}}{\sqrt{b_1 b_2}}L + O\left(\sqrt{b_1 b_2}\right)\right) < \mathcal{C}L\frac{rm}{\sqrt{b_1 b_2}}, \tag{B.69}$$

where $\mathcal{C}$ is a constant that is independent of the index $j$.

It remains to verify that the indicator functions $\mathbb{1}\left\{\left|\eta_j^{\leq L}\right| \geq \epsilon\sqrt{(\mathfrak{s}_\ell^2)^{\leq L}}\right\}$, from the lower truncated version of eq. (B.66), becomes zero when $n \to \infty$, which can be done by checking that the upper bound of $\left|\eta_j^{\leq L}\right|$ from eq. (B.69) in the limit gives a smaller value than the lower truncated version of $(\mathfrak{s}_\ell^2)^{\leq L}$ from eq. (B.65). This in turn can be done by dividing both of them with $\sqrt{\ell r m}$, and then compare their limits. Assuming that the threshold value $L$ is high enough to allow corollary B.19(c) to be used, i.e. that $(\mathfrak{s}_\ell^2)^{\leq L}$ and $\mathfrak{s}_\ell^2$ share

[9] Truncation arguments often requires the threshold value $L$ to go to $\infty$ in order for a conclusion to be obtained for the original expression, but this is not required for the present case under investigation (due to lemma B.8).

the same asymptotic expression, this becomes,

$$\frac{\left|\eta_j^{\leq L}\right|}{\sqrt{\ell r m}} \leq \mathcal{C} L \sqrt{\frac{m r}{\ell b_1 b_2}} \longrightarrow 0, \qquad \text{due to lemma C.3(c)}, \tag{B.70a}$$

$$\frac{\epsilon \sqrt{(\mathfrak{s}_\ell^2)^{\leq L}}}{\sqrt{\ell r m}} = \epsilon \cdot \sqrt{\frac{1}{m}\left\{\sum_{h=1}^{m} \boldsymbol{a}_h' W_{h:\boldsymbol{b}} \boldsymbol{a}_h + o(1)\right\}} \asymp \epsilon \cdot \sqrt{\frac{1}{m}\sum_{h=1}^{m} \boldsymbol{a}_h' W_{h:\boldsymbol{b}} \boldsymbol{a}_h}. \tag{B.70b}$$

Assumption 2.2(b) ensures that $W_{h:\boldsymbol{b}}$ (from definition B.10) converges to some non-zero matrix (as $h \to \infty$ and $\boldsymbol{b} \to \boldsymbol{0}^+$), and this implies that the limit of $\frac{1}{m}\sum_{h=1}^{m} \boldsymbol{a}_h' W_{h:\boldsymbol{b}} \boldsymbol{a}_h$ in eq. (B.70b) will be nonzero, from which it follows that the indicator function in eq. (B.66) becomes zero in the limit, i.e. that the Lindeberg condition is satisfied.

This implies that

$$\frac{\sum_{j=1}^{\ell} \eta_j^{\leq L}}{\sqrt{\mathfrak{s}_\ell^2}} \longrightarrow N(0,1), \tag{B.71}$$

which due to $\ell r \asymp n$ can be re-expressed as

$$n^{-1/2} \sum_{j=1}^{\ell} \eta_j^{\leq L} \longrightarrow \mathrm{N}\left(0, \sum_{h=1}^{m} \boldsymbol{a}_h' W_{h:\boldsymbol{b}} \boldsymbol{a}_h\right). \tag{B.72}$$

The proof of item (a) is now complete, since the four random variables $n^{-1/2}\wp_{\bar{m}}^{n}(\boldsymbol{a})$, $n^{-1/2}\wp_{\bar{m}}^{n|\leq L}(\boldsymbol{a})$, $n^{-1/2}(S_n^{(1)})^{\leq L}$ and $n^{-1/2}\sum_{j=1}^{\ell} \eta_j^{\leq L}$ all share the same limiting distribution (when $L$ is large enough).

The proof of item (b) follows from the Cramér-Wold theorem. □

The statements in theorem B.20 has to be interpreted as an approximate asymptotic distributions valid for large $m$ and $n$ and small $\boldsymbol{b}$. One part of the 'asymptotic problem' is the interpretation of an infinite-variate Gaussian distribution, but the main problem is the occurrence of the kernel function $K(\boldsymbol{w})$, which in the limit gives a degenerate Gaussian distribution in theorem B.20(b). This degeneracy in itself would not have been any issue if the target of interest had been the asymptotic behaviour of $n^{-1/2}\wp_{\bar{m}}^{n}$, but it requires some additional rescaling before the Klimko-Nelson approach in theorem B.1 can be used to investigate the asymptotic properties of the estimates $\widehat{\boldsymbol{\theta}}_{\bar{m}:n}$, see appendix B.3 for details.

**Corollary B.21.** *Given the same assumptions as in theorem B.20, the following asymptotic result holds true*

$$n^{-1/2}\sqrt{b_1 b_2}\,\boldsymbol{\nabla}_{\overline{m}} Q_{\overline{m}:n}(\boldsymbol{\theta}_{\overline{m}:\boldsymbol{b}}) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \bigoplus_{h=1}^{m} W_{h:\boldsymbol{b}}\right), \tag{B.73}$$

*i.e. asymptotically* $5m$*-variate normal.*

*Proof.* Lemma B.14 states that $\boldsymbol{\mathfrak{Q}}^n_{\overline{m}}$ and $\sqrt{b_1 b_2}\,\boldsymbol{\nabla}_{\overline{m}} Q_{\overline{m}:n}(\boldsymbol{\theta}_{\overline{m}:\boldsymbol{b}})$ have the same limiting distribution, and the result thus follows from theorem B.20(b). □

## B.3 The asymptotic results for $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}}$

The final details needed for the investigation of the asymptotic properties of $\widehat{f}^m_{\boldsymbol{v}}(\omega)$ will now be presented. (Confer appendix B.2.3 for an interpretation of the $m$ that occurs in the limiting distribution.)

**Theorem B.22.** *Under the same assumptions as in theorem B.20, the estimated parameter vector* $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}}$ *converges towards the true parameter vector* $\boldsymbol{\theta}_{\boldsymbol{v}|\overline{m}}$ *in the following manner.*

$$\sqrt{n(b_1 b_2)^3}\cdot\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}} - \boldsymbol{\theta}_{\boldsymbol{v}|\overline{m}}\right) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \Sigma_{\boldsymbol{v}|\overline{m}}\right), \tag{B.74}$$

*where* $\Sigma_{\boldsymbol{v}|\overline{m}} := \bigoplus_{h=1}^{m} \Sigma_{\boldsymbol{v}|h}$*, i.e.* $\Sigma_{\boldsymbol{v}|\overline{m}}$ *is the direct sum of the covariance matrices* $\Sigma_{\boldsymbol{v}|h}$ *that corresponds to* $\sqrt{n(b_1 b_2)^3}\cdot\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|h|\boldsymbol{b}} - \boldsymbol{\theta}_{\boldsymbol{v}|h}\right)$.

*Proof.* Under the given assumptions, corollary B.21 states that the fourth requirement of theorem B.1 (the Klimko-Nelson approach) holds true for the local penalty function $Q_{\overline{m}:n}\left(\boldsymbol{\theta}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}}\right)$ in the general case where $m \to \infty$ and $\boldsymbol{b} \to \mathbf{0}^+$ when $n \to \infty$. The three remaining requirements holds true by the same arguments that was used in appendix B.1.3, so the Klimko-Nelson approach can be used to obtain an asymptotic result for the difference of the estimate $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}}$ and the true parameter $\boldsymbol{\theta}_{\boldsymbol{v}|\overline{m}}$.

As in Tjøstheim and Hufthammer [2013], it will be instructive to first consider the simpler case where $m$ and $\boldsymbol{b}$ were fixed. In this case, the asymptotic result obtained from theorem B.1 takes the form,

$$\sqrt{n}\cdot\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}} - \boldsymbol{\theta}_{\boldsymbol{v}|\overline{m}}\right) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \Sigma_{\boldsymbol{v}|\overline{m}}\right), \tag{B.75}$$

with $\Sigma_{\boldsymbol{v}|\bar{m}} \coloneqq V_{\boldsymbol{v}|\bar{m}}^{-1} W_{\boldsymbol{v}|\bar{m}} V_{\boldsymbol{v}|\bar{m}}^{-1}$, where the $5m \times 5m$ matrices $V_{\boldsymbol{v}|\bar{m}}$ and $W_{\boldsymbol{v}|\bar{m}}$ can be represented as

$$V_{\boldsymbol{v}|\bar{m}} = \bigoplus_{h=1}^{m} V_{\boldsymbol{v}|h}, \qquad W_{\boldsymbol{v}|\bar{m}} = \bigoplus_{h=1}^{m} W_{\boldsymbol{v}|h}, \tag{B.76}$$

i.e. they are the direct sums of the $5 \times 5$ matrices $V_{\boldsymbol{v}|h}$ and $W_{\boldsymbol{v}|h}$ that corresponds to the bivariate penalty functions used for the investigation of the parameter vectors $\boldsymbol{\theta}_{\boldsymbol{v}|h|\boldsymbol{b}}$.

Since $V_{\boldsymbol{v}|\bar{m}}$ is the direct sum of the invertible matrices $V_{\boldsymbol{v}|h}$, it follows that $V_{\boldsymbol{v}|\bar{m}}^{-1}$ is the direct sum of $V_{\boldsymbol{v}|h}^{-1}$ (see e.g. Horn and Johnson [2012, p.31]). This implies that the matrix of interest can be expressed as $\Sigma_{\boldsymbol{v}|\bar{m}} = \bigoplus_{h=1}^{m} \Sigma_{\boldsymbol{v}|h}$, where $\Sigma_{\boldsymbol{v}|h} \coloneqq V_{\boldsymbol{v}|h}^{-1} W_{\boldsymbol{v}|h} V_{\boldsymbol{v}|h}^{-1}$ are the covariance matrices that corresponds to $\sqrt{n} \cdot \left( \widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|h|\boldsymbol{b}} - \boldsymbol{\theta}_{\boldsymbol{v}|h} \right)$, i.e. a bivariate result like the one in Tjøstheim and Hufthammer [2013, Th. 1].

For the general situation, when $m \to \infty$ and $\boldsymbol{b} \to \mathbf{0}^{+}$ when $n \to \infty$, it is necessary with an additional scaling in order to get a covariance matrix with finite entries. Obviously, a factor $\sqrt{b_1 b_2}$ must be included in order to balance the effect of the kernel function $K_{h:\boldsymbol{b}}$.

Moreover, since the limiting matrices of $V_{\boldsymbol{v}|h}$ and $W_{\boldsymbol{v}|h}$ turns out to have rank one, an additional scaling is required in order to obtain a covariance matrix with finite entries. This case is treated in Tjøstheim and Hufthammer [2013, Th. 3], from which it follows that the scaling factor must be $\sqrt{(b_1 b_2)^3}$. □

## B.4 An extension to two different points, i.e. both $\boldsymbol{v}$ and $\breve{\boldsymbol{v}}$

The previous analysis was restricted to the case where one point was used throughout, which is sufficient for the investigation of the asymptotic properties of the $m$-truncated estimates $\widehat{f}_{\boldsymbol{v}}^{m}(\omega)$ for a point $\boldsymbol{v}$ that lies upon the diagonal (see theorem 2.8) or for general points $\boldsymbol{v} \in \mathbb{R}^2$ when the time series under investigation is time reversible (see theorem 2.9).

An investigation of the $m$-truncated estimates $\widehat{f}_{\boldsymbol{v}}^{m}(\omega)$ for points $\boldsymbol{v} = (v_1, v_2)$ that lies off the diagonal, i.e. $v_1 \neq v_2$, requires some minor modifications of the setup leading to theorem B.22, as discussed in the proof of the following theorem.

**Theorem B.23.** *Consider the same setup as in theorem B.20, but with the modification that the point $\boldsymbol{v} = (v_1, v_2)$ lies off the diagonal, and with the added requirement that the bivariate densities $g_h(\boldsymbol{y}_h)$ does not possess diagonal symmetry. With $\breve{\boldsymbol{v}} = (v_2, v_1)$ the diagonal reflection of $\boldsymbol{v}$, the two parameter vectors $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\bar{m}|\boldsymbol{b}}$ and $\widehat{\boldsymbol{\theta}}_{\breve{\boldsymbol{v}}|\bar{m}|\boldsymbol{b}}$ can be combined to a vector*

$\widehat{\boldsymbol{\Theta}}_{\underline{\overline{m}}|\boldsymbol{b}}(\boldsymbol{v},\breve{\boldsymbol{v}}) = \left[\widehat{\boldsymbol{\theta}}'_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}}, \widehat{\boldsymbol{\theta}}'_{\breve{\boldsymbol{v}}|\overline{m}|\boldsymbol{b}}\right]'$, *possessing the following asymptotic behaviour.*

$$\sqrt{n(b_1 b_2)^3} \cdot \left(\widehat{\boldsymbol{\Theta}}_{\underline{\overline{m}}|\boldsymbol{b}}(\boldsymbol{v},\breve{\boldsymbol{v}}) - \boldsymbol{\Theta}_{\underline{\overline{m}}}(\boldsymbol{v},\breve{\boldsymbol{v}})\right) \xrightarrow{d} \mathrm{N}\left(\boldsymbol{0}, \begin{bmatrix} \Sigma_{\boldsymbol{v}|\overline{m}} & 0 \\ 0 & \Sigma_{\breve{\boldsymbol{v}}|\overline{m}} \end{bmatrix}\right), \tag{B.77}$$

*where the matrices* $\Sigma_{\boldsymbol{v}|\overline{m}}$ *and* $\Sigma_{\breve{\boldsymbol{v}}|\overline{m}}$ *are as given in theorem B.22.*

*Proof.* This result follows when the Klimko-Nelson approach is used upon the local penalty-function

$$Q_{\underline{\overline{m}}:n}\left(\boldsymbol{\Theta}_{\underline{\overline{m}}|\boldsymbol{b}}(\boldsymbol{v},\breve{\boldsymbol{v}})\right) \coloneqq Q_{\overline{m}:n}\left(\boldsymbol{\theta}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}}\right) + Q_{\overline{m}:n}\left(\boldsymbol{\theta}_{\breve{\boldsymbol{v}}|\overline{m}|\boldsymbol{b}}\right), \tag{B.78}$$

i.e. the four requirements in items *(A1)* to *(A4)* of theorem B.1 must be verified for this new penalty function. The function $Q_{\overline{m}:n}$ on the right side of eq. (B.78) is the penalty function encountered in the investigation of $\boldsymbol{\theta}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}}$, i.e. the same observations $\{Y_t\}_{t=1}^n$ occurs in both the first and second term, but the point of interest will be $\boldsymbol{v}$ in the first one and $\breve{\boldsymbol{v}}$ in the second one.

The requirement that $\boldsymbol{v}$ lies off the diagonal together with the requirement that none of the bivariate densities $g_h(\boldsymbol{y}_h)$ possess diagonal symmetry implies that different approximating local Gaussian densities occurs for the different points and different lags, so it can be assumed that there is no common parameters in $\boldsymbol{\theta}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}}$ and $\boldsymbol{\theta}_{\breve{\boldsymbol{v}}|\overline{m}|\boldsymbol{b}}$. This implies that the arguments used to verify the three first requirements of theorem B.1 for the penalty function $Q_{\overline{m}:n}$ (see lemmas B.3 to B.5), also will work upon the combined penalty function $Q_{\underline{\overline{m}}:n}$, and it will in particular be the case that the Hessian matrix $V_{\underline{\overline{m}}|\boldsymbol{b}:n}$ occurring in lemma B.4 can be written as the direct sum of the matrices that corresponds to $Q_{\overline{m}:n}\left(\boldsymbol{\theta}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}}\right)$ and $Q_{\overline{m}:n}\left(\boldsymbol{\theta}_{\breve{\boldsymbol{v}}|\overline{m}|\boldsymbol{b}}\right)$, i.e. $V_{\underline{\overline{m}}|\boldsymbol{b}}(\boldsymbol{v},\breve{\boldsymbol{v}}) = V_{\overline{m}|\boldsymbol{b}:n}(\boldsymbol{v}) \oplus V_{\overline{m}|\boldsymbol{b}:n}(\breve{\boldsymbol{v}})$, where the points of interest have been included in the notation to keep track of the components.

The investigation of the fourth requirement of the Klimko-Nelson approach for the penalty function $Q_{\underline{\overline{m}}:n}$ requires some minor modifications of the constructions that was encountered in appendix B.2.1. Both $X_{hq:t}^n(\boldsymbol{v})$ and $X_{hq:t}^n(\breve{\boldsymbol{v}})$ (for $h = 1,\ldots,m$ and $q = 1,\ldots,5$) are needed, and the final random variable will include both $\boldsymbol{v}$ and $\breve{\boldsymbol{v}}$ versions of the variables $Z_{hq:t}^n$, $\mathcal{Q}_{hq}^n$, $Z_{\overline{m}:t}^n(\boldsymbol{a})$, $\boldsymbol{Z}_{\overline{m}:t}^n$, $\mathcal{Q}_{\overline{m}}^n(\boldsymbol{a})$ and $\boldsymbol{\mathcal{Q}}_{\overline{m}}^n$.

A minor revision of lemma B.14 proves that the same limiting distribution occurs for the $\sqrt{b_1 b_2}$-scaled gradient of $Q_{\underline{\overline{m}}:n}\left(\boldsymbol{\Theta}_{\underline{\overline{m}}|\boldsymbol{b}}(\boldsymbol{v},\breve{\boldsymbol{v}})\right)$ and for the random variable $\boldsymbol{\mathcal{Q}}_{\underline{\overline{m}}}^n(\boldsymbol{v},\breve{\boldsymbol{v}}) \coloneqq [\boldsymbol{\mathcal{Q}}_{\overline{m}}^n(\boldsymbol{v})', \boldsymbol{\mathcal{Q}}_{\overline{m}}^n(\breve{\boldsymbol{v}})']'$, and it is easy to see that $Z_{\underline{\overline{m}}:t}^n(\boldsymbol{a}_1,\boldsymbol{a}_2;\boldsymbol{v},\breve{\boldsymbol{v}}) \coloneqq Z_{\overline{m}:t}^n(\boldsymbol{a}_1;\boldsymbol{v}) + Z_{\overline{m}:t}^n(\boldsymbol{a}_2;\breve{\boldsymbol{v}})$ must take the place of $Z_{\overline{m}:t}^n(\boldsymbol{a})$ in the existing proofs. The key ingredient for the asymptotic investigation of $Z_{\underline{\overline{m}}:t}^n(\boldsymbol{a}_1,\boldsymbol{a}_2;\boldsymbol{v},\breve{\boldsymbol{v}})$ is a simple extension of lemma B.15(c) such that

it also covers the 'cross-term' cases $\mathrm{E}\big[X^n_{hq:i}(\boldsymbol{v}) \cdot X^n_{jr:k}(\breve{\boldsymbol{v}})\big]$ and verifies that these cases are asymptotically negligible. This follows from the results stated in lemma C.7

The statement for $Z^n_{\overline{m}:t}(\boldsymbol{a})$ given in corollary B.17(b) extends trivially to the present case, since the asymptotic behaviour are unaffected by the adjustment that a sum of length $m$ is replaced by two sums of length $m$. The statement in lemma B.18 remains the same too, but some minor adjustments are needed in the proof: First of all, from the definition of $Z^n_{\underline{\overline{m}}:t}(\boldsymbol{a}_1, \boldsymbol{a}_2; \boldsymbol{v}, \breve{\boldsymbol{v}})$, it follows that

$$\begin{aligned} Z^n_{\underline{\overline{m}}:i}(\boldsymbol{a}_1, \boldsymbol{a}_2; \boldsymbol{v}, \breve{\boldsymbol{v}}) \cdot Z^n_{\underline{\overline{m}}:k}(\boldsymbol{a}_1, \boldsymbol{a}_2; \boldsymbol{v}, \breve{\boldsymbol{v}}) &= Z^n_{\overline{m}:i}(\boldsymbol{a}_1; \boldsymbol{v}) \cdot Z^n_{\overline{m}:k}(\boldsymbol{a}_1; \boldsymbol{v}) + Z^n_{\overline{m}:i}(\boldsymbol{a}_1; \boldsymbol{v}) \cdot Z^n_{\overline{m}:k}(\boldsymbol{a}_2; \breve{\boldsymbol{v}}) \\ &\quad + Z^n_{\overline{m}:k}(\boldsymbol{a}_1; \boldsymbol{v}) \cdot Z^n_{\overline{m}:i}(\boldsymbol{a}_2; \breve{\boldsymbol{v}}) + Z^n_{\overline{m}:i}(\boldsymbol{a}_2; \breve{\boldsymbol{v}}) \cdot Z^n_{\overline{m}:k}(\boldsymbol{a}_2; \breve{\boldsymbol{v}}), \end{aligned} \tag{B.79}$$

and only the parts that contains both $\boldsymbol{v}$ and $\breve{\boldsymbol{v}}$ needs to be investigated (since the other terms already are covered by the existing results). The statement that must be verified reduces to

$$\frac{1}{n} \sum_{\substack{i,k=1 \\ i \neq k}}^{n} |Z^n_{\overline{m}:i}(\boldsymbol{a}_1; \boldsymbol{v}) \cdot Z^n_{\overline{m}:k}(\boldsymbol{a}_2; \breve{\boldsymbol{v}})| = o(1)\,, \tag{B.80}$$

and it is straightforward to verify that this sum can be realised as

$$\sum_{\ell=1}^{n-1} \left(1 - \frac{\ell}{n}\right) I^n_{\overline{m}:\ell}(\boldsymbol{a}_1, \boldsymbol{a}_2; \boldsymbol{v}, \breve{\boldsymbol{v}}) + \sum_{\ell=1}^{n-1} \left(1 - \frac{\ell}{n}\right) I^n_{\overline{m}:\ell}(\boldsymbol{a}_2, \boldsymbol{a}_1; \breve{\boldsymbol{v}}, \boldsymbol{v}), \tag{B.81}$$

where $I^n_{\overline{m}:\ell}(\boldsymbol{a}_1, \boldsymbol{a}_2; \boldsymbol{v}, \breve{\boldsymbol{v}}) \coloneqq |\mathrm{E}[Z^n_{\overline{m}:0}(\boldsymbol{a}_1, \boldsymbol{v}) \cdot Z^n_{\overline{m}:\ell}(\boldsymbol{a}_2, \breve{\boldsymbol{v}})]|$, with $I^n_{\overline{m}:\ell}(\boldsymbol{a}_2, \boldsymbol{a}_1; \breve{\boldsymbol{v}}, \boldsymbol{v})$ defined in the obvious manner by interchanging the parameters and the points. The desired result follows from this, since the remaining part of the proof of lemma B.18 (using the adjusted version of lemma B.15(c)) gives that the two sums in eq. (B.81) both are $o(1)$.

The investigation of the variance of $Z^n_{\underline{\overline{m}}:t}(\boldsymbol{a}_1, \boldsymbol{a}_2; \boldsymbol{v}, \breve{\boldsymbol{v}})$ is straight forward, i.e. the standard formula for the variance of a sum of random variables gives

$$\mathrm{Var}\big(Z^n_{\underline{\overline{m}}:t}(\boldsymbol{a}_1, \boldsymbol{a}_2; \boldsymbol{v}, \breve{\boldsymbol{v}})\big) = \mathrm{Var}(Z^n_{\overline{m}:t}(\boldsymbol{a}_1, \boldsymbol{v})) + 2\,\mathrm{Cov}(Z^n_{\overline{m}:t}(\boldsymbol{a}_1, \boldsymbol{v}), Z^n_{\overline{m}:t}(\boldsymbol{a}_2, \breve{\boldsymbol{v}})) + \mathrm{Var}(Z^n_{\overline{m}:t}(\boldsymbol{a}_2, \breve{\boldsymbol{v}}))\,,$$

and the revised version of lemma B.15(c) implies that the covariance part of this expression is asymptotically negligible. The two variances are already covered by the existing version of corollary B.19(a), and from this it is clear that the asymptotically non-negligible parts

can be written as

$$\boldsymbol{a}'_{\underline{\overline{m}}} \cdot W_{\underline{\overline{m}}:\boldsymbol{b}} \cdot \boldsymbol{a}_{\underline{\overline{m}}} \coloneqq [\boldsymbol{a}'_1, \boldsymbol{a}'_2] \cdot (W_{\overline{m}:\boldsymbol{b}}(\boldsymbol{v}) \oplus W_{\overline{m}:\boldsymbol{b}}(\breve{\boldsymbol{v}})) \cdot \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \end{bmatrix} = \boldsymbol{a}'_1 \cdot W_{\overline{m}:\boldsymbol{b}}(\boldsymbol{v}) \cdot \boldsymbol{a}_1 + \boldsymbol{a}'_2 \cdot W_{\overline{m}:\boldsymbol{b}}(\breve{\boldsymbol{v}}) \cdot \boldsymbol{a}_2, \tag{B.82}$$

whereas the asymptotically negligible parts of corollary B.19(a) remains as before. This is sufficient for the revision of corollary B.19 (since items (b) and (c) follows from item (a) and lemma B.18)

Finally, theorem B.20 can now be updated based on the matrix $W_{\underline{\overline{m}}:\boldsymbol{b}} \coloneqq W_{\overline{m}:\boldsymbol{b}}(\boldsymbol{v}) \oplus W_{\overline{m}:\boldsymbol{b}}(\breve{\boldsymbol{v}})$, and with some minor adjustments of the proof, i.e. new cross-terms are asymptotically negligible and sums of length $m$ are replaced with two sums of length $m$, it follows that

$$n^{-1/2}\, Q_{\underline{\overline{m}}:n}\big(\boldsymbol{\Theta}_{\underline{\overline{m}}|\boldsymbol{b}}(\boldsymbol{v}, \breve{\boldsymbol{v}})\big) \xrightarrow{d} \mathrm{N}(\boldsymbol{0}, W_{\overline{m}:\boldsymbol{b}}(\boldsymbol{v}) \oplus W_{\overline{m}:\boldsymbol{b}}(\breve{\boldsymbol{v}}))\,. \tag{B.83}$$

The revised version of corollary B.21 is as before trivial to prove, which completes the investigation of the fourth requirement needed in order to use the Klimko-Nelson approach. Basic linear algebra together with theorem B.22 now finishes the proof. □

The arguments above could (under suitable assumptions) have been formulated in a more general setup, leading to a result that shows that the parameter vectors $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}_i|\overline{m}|\boldsymbol{b}}$ corresponding to different points $\{\boldsymbol{v}_i\}_{i=1}^{\nu}$ will be jointly asymptotically normal and pairwise asymptotically independent. The asymptotically independent property are inherited by the corresponding estimated local Gaussian spectral densities $\widehat{f}^{\,m}_{\boldsymbol{v}_i}(\omega)$, and this enables an alternative smoothing strategy for the estimated local Gaussian spectral densities at a given point $\boldsymbol{v}$, see appendix B.5. However, the added computational cost incurred by such an estimation approach may make this a less interesting topic of investigation.

## B.5 An alternative smoothing strategy?

The previously defined estimates $\widehat{f}^{\,m}_{\boldsymbol{v}}(\omega)$ of $f_{\boldsymbol{v}}(\omega)$ was based on a weighting function $\lambda_m(h)$ that worked upon the estimated values $\widehat{\rho}_{\boldsymbol{v}}(h)$, but it should for the record be noted that an alternative approach could have been applied too.

The point is that it is possible to extend the result of appendix B.4 to show that the estimated $m$-truncated local Gaussian spectral densities $\widehat{f}^{\,m}_{\boldsymbol{v}_i}(\omega)$ corresponding to different points $\{\boldsymbol{v}_i\}_{i=1}^{\nu}$ will be jointly asymptotically normal and pairwise asymptotically independent (when $m \to \infty$ and $\boldsymbol{b} \to \boldsymbol{0}^+$ as $n \to \infty$). This enables an alternative smoothing

strategy, where an estimate $\widehat{f}^{m}_{\boldsymbol{v}}(\omega)$ for a given point $\boldsymbol{v}$ could be based on a weighting of the values of $\widehat{f}^{m}_{\boldsymbol{v}_i}(\omega)$ in a grid of points surrounding $\boldsymbol{v}$.

This alternative approach shares some superficial similarities with the one used when the ordinary global spectrum $f(\omega)$ is computed based on the periodogram, see e.g. Brockwell and Davis [1986] for details. However, the efficiency of the periodogram-approach in the estimation of $f(\omega)$ is due to the *Fast Fourier Transform*, which implies that the periodogram can be computed directly from the observations without the need for an explicit computation of all of the estimated autocovariances $\widehat{\rho}(h)$, and that shortcut is not available for the local Gaussian case. The computational load would thus become much larger for the local Gaussian case if such an averaging-approach was applied.

# Appendix C: Technical details

This section collects some technical details that would have impeded the flow of the main argument if they had been included throughout the paper. A brief overview: Appendix C.1 discuss the *diagonal folding property* of the local Gaussian autocorrelations $\rho_{\boldsymbol{v}}(h)$ and appendix C.2 considers the special case of time-reversible time series. Appendix C.3 collects technical results related to the asymptotic relationship between $n$, $m$ and $\boldsymbol{b}$, whereas appendix C.4 shows that the assumptions on the kernel function $K(\boldsymbol{w})$ and the score functions $u_{hq:\boldsymbol{b}}(\boldsymbol{w})$ implies that some integrals are finite (which implies that assumption 2.1(g) will be trivially satisfied if the bivariate densities $g_h(\boldsymbol{y}_h)$ are finite). Appendix C.5 contains a few basic definitions/comments related to $\alpha$-mixing, $\sigma$-algebras and $L^{\nu}$-spaces.

## C.1 The diagonal folding property of $\rho_{\boldsymbol{v}}(h)$

The following simple observation about $\rho_{\boldsymbol{v}}(h)$ is of interest both for theoretical and computational aspects of the local Gaussian spectral density $f_{\boldsymbol{v}}(\omega)$.

**Lemma C.1.** *For a strictly stationary time series $\{Y_t\}_{t\in\mathbb{Z}}$ and a point $\boldsymbol{v} = (v_1, v_2)$, the following symmetry property (diagonal folding) holds for the local Gaussian autocorrelation,*

$$\rho_{\boldsymbol{v}}(-h) = \rho_{\breve{\boldsymbol{v}}}(h), \tag{C.1}$$

*where $\breve{\boldsymbol{v}} = (v_2, v_1)$ is the diagonal reflection of $\boldsymbol{v}$.*

*Proof.* This is a simple consequence of the symmetrical nature of the bivariate random variables $\boldsymbol{Y}_{h:t} \coloneqq (Y_h, Y_0)$ and $\boldsymbol{Y}_{-h:t} \coloneqq \big(Y_{-h}, Y_0\big)$, which due to the connection between the corresponding cumulative density functions

$$\begin{aligned} G_{-h}\big(y_{-h}, y_0\big) &= \mathrm{P}\big(Y_{-h} \leq y_{-h}, Y_0 \leq y_0\big) = \mathrm{P}\big(Y_0 \leq y_0, Y_{-h} \leq y_{-h}\big) = \mathrm{P}\big(Y_h \leq y_0, Y_0 \leq y_{-h}\big) \\ &= G_h\big(y_0, y_{-h}\big) \end{aligned} \tag{C.2}$$

gives the following property[10] for the probability density functions,

$$g_{-h}(y_{-h}, y_0) = g_h(y_0, y_{-h}) \,. \tag{C.3}$$

This implies that $g_{-h}(\boldsymbol{v}) = g_h(\breve{\boldsymbol{v}})$, and the symmetry does moreover induce a symmetrical relation between the parameters $\boldsymbol{\theta}_{-h}(\boldsymbol{v})$ of the local Gaussian approximation of $g_{-h}$ at $\boldsymbol{v}$ and the parameters $\boldsymbol{\theta}_h(\breve{\boldsymbol{v}})$ of the local Gaussian approximation of $g_h$ at $\breve{\boldsymbol{v}}$, i.e. if $\boldsymbol{\theta}_{-h}(\boldsymbol{v}) = [\mu_1, \mu_2, \sigma_{11}, \sigma_{22}, \rho]'$ then $\boldsymbol{\theta}_h(\breve{\boldsymbol{v}}) = [\mu_2, \mu_1, \sigma_{22}, \sigma_{11}, \rho]'$. Equation (C.1) follows since $\rho$ in these two vectors respectively represents $\rho_{\boldsymbol{v}}(-h)$ and $\rho_{\breve{\boldsymbol{v}}}(h)$, and this completes the proof. □

A trivial consequence of the *diagonal folding property* in lemma C.1 is that the local Gaussian autocorrelation becomes an even function of the lag $h$ when $v_1 = v_2$.

## C.2 Time-reversible time series

Additional symmetry properties are present for time reversible time series, which implies that the local Gaussian spectral densities $f_{\boldsymbol{v}}(\omega)$ always are real-valued for such time series, see definition 2.2 and theorem 2.9.

The following simple result follows immediately from definition 2.2.

**Lemma C.2.** *If* $\{Y_t\}_{t\in\mathbb{Z}}$ *is time reversible, then*

$$g_h(v_1, v_2) = g_h(v_2, v_1) \tag{C.4}$$

*for all points* $\boldsymbol{v} = (v_1, v_2) \in \mathbb{R}^2$ *and all* $h \in \mathbb{N}$*, which implies*

$$\rho_{\boldsymbol{v}}(-h) = \rho_{\boldsymbol{v}}(h). \tag{C.5}$$

*Proof.* The time reversibility of $\{Y_t\}_{t\in\mathbb{Z}}$ implies that $(Y_h, Y_0)$ and $(Y_{-h}, Y_0)$ have the same joint distribution, i.e.

$$G_{-h}(y_{-h}, y_0) = \mathrm{P}(Y_{-h} \le y_{-h}, Y_0 \le y_0) = \mathrm{P}(Y_h \le y_{-h}, Y_0 \le y_0) = G_h(y_{-h}, y_0) \,.$$

Together with the observation in eq. (C.2), this gives the diagonal symmetry stated in eq. (C.4). The statement for the local Gaussian autocorrelations follows by the same reasoning as in the proof of lemma C.1. □

[10] This must not be confused with the property that $g_h$ and $g_{-h}$ themselves are symmetric around the diagonal, for that will in general not be the case.

## C.3 Two limit theorems — and one comment

This section contains two lemmas and one comment. Lemma C.3 combines a check of the internal consistency of assumption 2.3 with the limits needed for the small block-large block argument in theorem B.20, whereas lemma C.4 takes care of the two limits needed in order to prove that the *off the diagonal components* in lemma B.18 are asymptotically negligible. The comment at the end of this section has been included due to the remark at the end of section 2.4.3 in the main document.

**Lemma C.3.** *Under assumption 2.3, the following holds.*

*(a) There exists integers $s$ that makes items (e) and (f) of assumption 2.3 compatible.*

*(b) There exists integers $s$ and constants $c \coloneqq c_n \to \infty$, such that*

$$c \cdot s = o\Big(\sqrt{nb_1b_2/m}\Big)\,, \qquad \sqrt{nm/b_1b_2}\cdot c\cdot \alpha(s-m+1) \longrightarrow 0. \tag{C.6}$$

*(c) There exists integers $s$ and constants $c$, such that with $r$, $\ell$ and $\vartheta$ given as the integers*

$$r = r_n \coloneqq \left\lfloor \frac{\sqrt{nb_1b_2/m}}{c} \right\rfloor, \qquad \ell = \ell_n \coloneqq \left\lfloor \frac{n}{r+s} \right\rfloor, \qquad \vartheta = \vartheta_n \coloneqq s-m+1, \tag{C.7}$$

*the following limits occur when $n \to \infty$:*

$$\frac{s}{r} \longrightarrow 0; \qquad \ell\alpha(\vartheta) \longrightarrow 0; \qquad \frac{mr}{n} \longrightarrow 0; \qquad \frac{mr}{\ell b_1 b_2} \longrightarrow 0; \qquad \frac{m\ell s}{n} \longrightarrow 0. \tag{C.8}$$

*Proof.* Item (a) will be established by first observing that it is possible to find integers $s$ that ensures that assumption 2.3(f) is compatible with the requirement $m = o\big((nb_1b_2)^{\xi}\big)$, for any $\xi \in \big(0, \frac{1}{3}\big)$, and then checking that the exponent $\tau/(2+5\tau) - \lambda$ lies in this interval.

Observe that it is impossible to have $m = o(s)$ and $s = o\Big(\sqrt{nb_1b_2/m}\Big)$ when $m \geq \sqrt{nb_1b_2/m}$, which implies $m < \sqrt{nb_1b_2/m}$, which is equivalent to $m < (nb_1b_2)^{1/3}$. Some extra leeway is needed in order to construct the desired integers $s$, so consider the requirement

$$m = o\big((nb_1b_2)^{1/3-\zeta}\big)\,, \qquad \text{for some } \zeta \in \big(0, \tfrac{1}{3}\big)\,. \tag{C.9}$$

Define the integers $s$ by $s \coloneqq m \cdot \mathfrak{s}$, where $\mathfrak{s} \coloneqq 1 \vee \big\lfloor (nb_1b_2)^{\zeta/2} \big\rfloor$, and note that this construction ensures that $s$ goes to $\infty$. Further, $m = o(s)$ holds since $m/s = 1/\mathfrak{s} \to 0$, and $s = o\Big(\sqrt{nb_1b_2/m}\Big)$

holds since

$$\frac{s}{\sqrt{nb_1b_2/m}} \asymp \frac{m\cdot(nb_1b_2)^{\zeta/2}}{(nb_1b_2/m)^{1/2}} = \frac{m^{3/2}}{(nb_1b_2)^{(1-\zeta)/2}} = \left[\frac{m}{(nb_1b_2)^{(1-\zeta)/3}}\right]^{3/2}$$
$$= \left[\frac{1}{(nb_1b_2)^{2\zeta/3}}\cdot\frac{m}{(nb_1b_2)^{1/3-\zeta}}\right]^{3/2} \to \left[\frac{1}{\infty}\cdot 0\right]^{3/2} = 0. \tag{C.10}$$

This implies that the desired integers $s$ can be found whenever $m = o\big((nb_1b_2)^{\xi}\big)$, with $\xi \in \big(0, \frac{1}{3}\big)$. Since the value of $\tau/(2+5\tau) - \lambda$ lies in the interval $\big(0, \frac{1}{5}\big)$, the proof of item (a) is complete.

For items (b) and (c), the integers $s$ and constants $c$ can e.g. be defined as

$$s = 1 \vee \left\lfloor \left(\sqrt{nb_1b_2/m}\right)^{1-\eta} \right\rfloor, \qquad c = \left(\sqrt{nb_1b_2/m}\right)^{\eta/2}, \quad \text{for some } \eta \in (0,1). \tag{C.11}$$

Since $1-\eta$ and $\eta/2$ are in $(0,1)$, it follows from assumption 2.3(b) that $s$ and $c$ goes to $\infty$ as required. A quick inspection reveals that the product $c\cdot s$ is $o\Big(\sqrt{nb_1b_2/m}\Big)$, proving the first part of eq. (C.6). For the second part of eq. (C.6), keep in mind the similarity with assumption 2.3(d), and observe that $c$ in the limit is asymptotically equivalent to $s^{\eta/2(1-\eta)}$. Since $\eta$ can be selected such that the exponent $\eta/2(1-\eta)$ becomes smaller than any $\tau > 0$, the second statement holds too, which completes the proof of item (b).

In order to prove item (c), note that a floor-function $\lfloor x \rfloor$ in a denominator can be ignored in the limit $x \to \infty$, since $x \asymp \lfloor x \rfloor$, that is $\lim x/\lfloor x \rfloor = 1$. Moreover, observe that assumption 2.3(b) implies that $n/m$ goes to $\infty$. With these observations, all except the last limit in eq. (C.8) are trivial to prove, i.e.

$$\frac{s}{r} \asymp \frac{s}{\frac{\sqrt{nb_1b_2/m}}{c}} = \frac{c\cdot s}{\sqrt{nb_1b_2/m}} \to 0, \tag{C.12a}$$

$$\ell\alpha(\vartheta) \le \frac{n}{r+s}\alpha(\vartheta) \asymp \frac{n}{r}\alpha(\vartheta) \asymp \frac{n}{\frac{\sqrt{nb_1b_2/m}}{c}}\alpha(\vartheta) = \sqrt{nm/b_1b_2}\cdot c\cdot\alpha(\vartheta) \to 0, \tag{C.12b}$$

$$\frac{mr}{n} \le \frac{\frac{\sqrt{nb_1b_2/m}}{c}}{n/m} = \frac{\sqrt{b_1b_2}}{c\sqrt{n/m}} \to \frac{0}{\infty\cdot\infty} = 0, \tag{C.12c}$$

$$\frac{mr}{\ell b_1b_2} \asymp \frac{mr}{\frac{n}{r+s}b_1b_2} = \frac{r(r+s)}{nb_1b_2/m} \asymp \frac{r^2}{nb_1b_2/m} \le \frac{\frac{nb_1b_2/m}{c^2}}{nb_1b_2/m} = \frac{1}{c^2} \to 0. \tag{C.12d}$$

For the proof of $m\ell s/n \to 0$, the explicit expressions for $s$ and $c$ from eq. (C.11) will be needed, i.e.

$$\begin{aligned}\frac{m\ell s}{n} &\leq \frac{m\frac{n}{r+s}s}{n} = m\frac{s}{r+s} \asymp m\frac{s}{r} \asymp m\frac{c\cdot s}{\sqrt{nb_1b_2/m}} \leq m\frac{\left(\sqrt{nb_1b_2/m}\right)^{1-\eta/2}}{\sqrt{nb_1b_2/m}} \\ &= \frac{m}{(nb_1b_2/m)^{\eta/4}} = \frac{m^{1+\eta/4}}{(nb_1b_2)^{\eta/4}} = \left(\frac{m}{(nb_1b_2)^{\eta/(4+\eta)}}\right)^{(4+\eta)/4}. \end{aligned} \tag{C.13}$$

Assumption 2.3(e) states that $m = o\big((nb_1b_2)^{\tau/(2+5\tau)-\lambda}\big)$, and it is consequently sufficient to show that an $\eta$ can be found which gives $\tau/(2+5\tau)-\lambda \leq p(\eta) \coloneqq \eta/(4+\eta)$. Since $p'(\eta) = 4/(4+\eta)^2 > 0$, the highest value of $p(\eta)$ will be found at the upper end of the interval of available arguments. From the proof of item (b) it is known that $\eta/2(1-\eta) < \tau$, which gives the requirement $\eta < 2\tau/(1+2\tau)$. The value of $p(\eta)$ at the upper end of this interval is $\tau/(2+5\tau)$, and since $\lambda > 0$ it is possible to find an $\eta$ that satisfies $\tau/(2+5\tau)-\lambda \leq p(\eta) < \tau/(2+5\tau)$, which concludes the proof. □

**Lemma C.4.** *Under assumption 2.3, the sequence of integers defined by* $k_n+1 \coloneqq \lceil m^{2/a}\cdot|b_1b_2|^{(2-\nu)/a\nu}\rceil$ *satisfies the following two limit requirements.*

*(a)* $k_n \longrightarrow \infty$.

*(b)* $k_n m^2 b_1 b_2 \longrightarrow 0$.

*Proof.* The key requirements $\nu > 2$ and $a > 1-2/\nu$ (inherited from assumption 2.1(b)) ensures that $2/a > 0$ and $(2-\nu)/a\nu < 0$. As $m \to \infty$ and $\boldsymbol{b} \to \mathbf{0}^+$ when $n \to \infty$, it follows that $k_n \to \infty$, which proves item (a).

For item (b), observe that $k_n = \lceil m^{2/a}\cdot|b_1b_2|^{(2-\nu)/a\nu}\rceil - 1 < m^{2/a}\cdot|b_1b_2|^{(2-\nu)/a\nu}$ implies

$$\begin{aligned} k_n m^2 b_1 b_2 &< \left(m^{2/a}\cdot|b_1b_2|^{(2-\nu)/a\nu}\right)\cdot m^2 b_1 b_2 && \text{(C.14a)}\\ &= m^{2(1+1/a)}\cdot|b_1b_2|^{1+(2-\nu)/a\nu} && \text{(C.14b)}\\ &\leq m^{2(1+1/a)}\cdot\left|(b_1\vee b_2)^2\right|^{1+(2-\nu)/a\nu} && \text{(C.14c)}\\ &= \left\{m^{\{1+1/a\}/\{1+(2-\nu)/a\nu\}}\cdot(b_1\vee b_2)\right\}^{2(1+(2-\nu)/a\nu)} && \text{(C.14d)}\\ &= \left\{m^{\{\nu(a+1)\}/\{\nu(a-1)+2\}}\cdot(b_1\vee b_2)\right\}^{2(1+(2-\nu)/a\nu)}. && \text{(C.14e)}\end{aligned}$$

An inspection of the outermost exponent reveals

$$2\cdot\left(1+\frac{(2-\nu)}{a\nu}\right) = 2\cdot\frac{a-(1-2/\nu)}{a} > 0, \tag{C.15}$$

which together with assumption 2.3(c) concludes the proof of item (b). □

**A comment related to the remark at the end of section 2.4.3:** It is not required for the theoretical investigation, but it might still be of interest to mention the following observation: Consider a combination of a given point $\boldsymbol{v}$, a small bandwidth vector $\boldsymbol{b} = (b_1, b_2)$, and a *large* sample of size $n$ from a univariate time series $\{Y_t\}_{t\in\mathbb{Z}}$ that satisfies assumption 2.1. The number of lag-$h$ pairs in the vicinity of $\boldsymbol{v}$ will then, for each $h = 1, \ldots, m$, be of order $nb_1b_2 \cdot g_h(\boldsymbol{v})$ — and this will, when $g_h(\boldsymbol{v}) > 0$, go to infinity when $n \to \infty$ and $\boldsymbol{b} \to \boldsymbol{0}^+$.

Only a sketch of the argument will be given here, since the asymptotic theory does not build upon this observation: First select a $\boldsymbol{b}$-dependent region $\mathcal{V}_{\boldsymbol{b}}(\boldsymbol{v})$ around $\boldsymbol{v}$ to be the '$\boldsymbol{b}$-vicinity of $\boldsymbol{v}$', i.e. $\mathcal{V}_{\boldsymbol{b}}(\boldsymbol{v})$ should shrink when $\boldsymbol{b} \to \boldsymbol{0}^+$. The area of $\mathcal{V}_{\boldsymbol{b}}(\boldsymbol{v})$ should be given by some constant $\mathcal{A}$ times $b_1b_2$. From a sample of size $n$ there will be a total of $n-h$ lag-$h$ pairs, and the expected number of those in the region $\mathcal{V}_{\boldsymbol{b}}(\boldsymbol{v})$ will be $(n-h) \cdot \iint_{\mathcal{V}_{\boldsymbol{b}}(\boldsymbol{v})} g_h(\boldsymbol{y}_h)\,\mathrm{d}\boldsymbol{y}_h$. Assumption 2.1(d) implies that the bivariate density functions $g_h(\boldsymbol{y}_h)$ are continuous at $\boldsymbol{v}$, and it is thus clear that both $\inf_{\boldsymbol{y}_h\in\mathcal{V}_{\boldsymbol{b}}(\boldsymbol{v})} g_h(\boldsymbol{y}_h)$ and $\sup_{\boldsymbol{y}_h\in\mathcal{V}_{\boldsymbol{b}}(\boldsymbol{v})} g_h(\boldsymbol{y}_h)$ go to $g_h(\boldsymbol{v})$ when $\boldsymbol{b} \to \boldsymbol{0}^+$. The integral $\iint_{\mathcal{V}_{\boldsymbol{b}}(\boldsymbol{v})} g_h(\boldsymbol{y}_h)\,\mathrm{d}\boldsymbol{y}_h$ will thus be of order $\mathcal{A} \cdot b_1b_2 \cdot g_h(\boldsymbol{v})$ when $\boldsymbol{b} \to \boldsymbol{0}^+$, and the result follows.

This shows why it even for rather large samples might be hard to obtain good estimates of the local Gaussian spectral densities in the tails, where the densities $g_h(\boldsymbol{v})$ are low.

## C.4 Integrals based on the kernel and the score functions

The asymptotic properties of the random variables introduced in definitions B.11 to B.13 does of course depend upon the properties of the time series $\{Y_t\}_{t\in\mathbb{Z}}$ upon which they have been defined, but quite a few of the required properties does in fact only depend upon $K(\boldsymbol{w})$ and $u_{hq:\boldsymbol{b}}(\boldsymbol{w})$. Note that the treatment in this section exploits the property that the functions $u_{hq:\boldsymbol{b}}(\boldsymbol{w})$ all are quadratic polynomials in the variables $w_1$ and $w_2$, which implies that the inequalities from lemma C.5 is sufficient for the proofs of the asymptotic results given in lemma C.6.

**Lemma C.5.** *For $K(\boldsymbol{w})$ from definition B.9 (page 16), and $\nu > 2$ from assumption 2.1(b), the following holds:*

*(a)* $\left|\int_{\mathbb{R}^2} K(w_1, w_2)\, w_1^k w_2^\ell \,\mathrm{d}w_1\mathrm{d}w_2\right| < \infty, \qquad k, \ell \geq 0$ *and* $k + \ell \leq 5.$

*(b)* $\left|\int_{\mathbb{R}^2} K(w_1, w_2)^2 w_1^k w_2^\ell \,\mathrm{d}w_1\mathrm{d}w_2\right| < \infty, \qquad k, \ell \geq 0$ *and* $k + \ell \leq 5.$

*(c)* $K(w_1, w_2)\, w_1^k w_2^\ell \in L^\nu, \qquad k, \ell \geq 0$ *and* $k + \ell \leq 2.$

*Proof.* Since the kernel function by definition is non-negative, it follows that

$$\left| \int_{\mathbb{R}^2} K(w_1, w_2)\, w_1^k w_2^\ell \, \mathrm{d}w_1 \mathrm{d}w_2 \right| \leq \int_{\mathbb{R}^2} K(w_1, w_2)\, |w_1^k w_2^\ell| \, \mathrm{d}w_1 \mathrm{d}w_2, \tag{C.16}$$

which proves item (a), since eq. (B.27d) of definition B.9 implies that this is finite for the specified range of $k$ and $\ell$.

Since the kernel function is bounded, there is some constant $\mathcal{C}$ such that $K(\boldsymbol{w}) \leq \mathcal{C}$, which implies that

$$\left| \int_{\mathbb{R}^2} K(w_1, w_2)^2 w_1^k w_2^\ell \, \mathrm{d}w_1 \mathrm{d}w_2 \right| \leq \mathcal{C} \left| \int_{\mathbb{R}^2} K(w_1, w_2)\, w_1^k w_2^\ell \, \mathrm{d}w_1 \mathrm{d}w_2 \right|, \tag{C.17}$$

which due to item (a) is finite, thus item (b) holds true.

Next, note that $|K(w_1, w_2)\, w_1^k w_2^\ell|^\nu = |K(w_1, w_2)|^{(\nu-1)}\, |K(w_1, w_2)|\, |w_1^k w_2^\ell|^\nu \leq \mathcal{C}^{(\nu-1)} K(w_1, w_2)\, |w_1^k w_2^\ell|^\nu$, which gives the following inequality,

$$\left( \int_{\mathbb{R}^2} |K(w_1, w_2)\, w_1^k w_2^\ell|^\nu \, \mathrm{d}w_1 \mathrm{d}w_2 \right)^{1/\nu} \leq \mathcal{C}^{(\nu-1)/\nu} \left( \int_{\mathbb{R}^2} K(w_1, w_2)\, |w_1^k w_2^\ell|^\nu \, \mathrm{d}w_1 \mathrm{d}w_2 \right)^{1/\nu}, \tag{C.18}$$

from which it is clear that a proof of the finiteness of the right hand side of eq. (C.18) will imply item (c). Since the region of integration can be divided into $\mathcal{A}_{k\ell} = \{\boldsymbol{w} : |w_1^k w_2^\ell| \leq 1\}$ and $\mathcal{A}_{k\ell}^c = \mathbb{R}^2 \setminus \mathcal{A}_{k\ell}$, it follows from the non-negativeness of $K(\boldsymbol{w})$, and eqs. (B.27a) and (B.27d) of definition B.9, that

$$\int_{\mathcal{A}_{k\ell}} K(w_1, w_2)\, |w_1^k w_2^\ell|^\nu \, \mathrm{d}w_1 \mathrm{d}w_2 \leq \int_{\mathcal{A}_{k\ell}} K(w_1, w_2) \, \mathrm{d}w_1 \mathrm{d}w_2 \leq \int_{\mathbb{R}^2} K(w_1, w_2) \, \mathrm{d}w_1 \mathrm{d}w_2 = 1, \tag{C.19a}$$

$$\begin{aligned} \int_{\mathcal{A}_{k\ell}^c} K(w_1, w_2)\, |w_1^k w_2^\ell|^\nu \, \mathrm{d}w_1 \mathrm{d}w_2 &\leq \int_{\mathcal{A}_{k\ell}^c} K(w_1, w_2)\, |w_1^k w_2^\ell|^{\lceil \nu \rceil} \, \mathrm{d}w_1 \mathrm{d}w_2 \\ &\leq \int_{\mathbb{R}^2} K(w_1, w_2)\, |w_1^{k\lceil \nu \rceil} w_2^{\ell \lceil \nu \rceil}| \, \mathrm{d}w_1 \mathrm{d}w_2 < \infty, \end{aligned} \tag{C.19b}$$

where the last inequality follows since the assumption $k + \ell \leq 2$ ensures that $k\,\lceil \nu \rceil + \ell\,\lceil \nu \rceil \leq 2\,\lceil \nu \rceil$. The expression in eq. (C.18) is thus finite — and, as stated in item (c), $K(w_1, w_2)\, w_1^k w_2^\ell \in L^\nu$. □

**Lemma C.6.** *The following holds for $u_{hq:\boldsymbol{b}}(\boldsymbol{w})$ and $K_{h:\boldsymbol{b}}(\boldsymbol{y}_h - \boldsymbol{v})$ from definitions B.7 and B.9, and $\nu > 2$ from assumption 2.1(b):*

*(a)* $\int_{\mathbb{R}^2} \sqrt{b_1 b_2} K_{h:\boldsymbol{b}}(\boldsymbol{\zeta} - \boldsymbol{v})\, u_{hq:\boldsymbol{b}}(\boldsymbol{\zeta})\, \mathrm{d}\boldsymbol{\zeta} = O\big(\sqrt{b_1 b_2}\big)$.

*(b)* $\left(\int_{\mathbb{R}^2} \left|\sqrt{b_1 b_2} K_{h:\boldsymbol{b}}(\boldsymbol{\zeta} - \boldsymbol{v})\, u_{hq:\boldsymbol{b}}(\boldsymbol{\zeta})\right|^{\nu} \mathrm{d}\boldsymbol{\zeta}\right)^{1/\nu} = O\big(|b_1 b_2|^{(2-\nu)/2\nu}\big)$.

*(c)* *Let* $\mathcal{K}_{qr,hj:\boldsymbol{b}}(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2) \coloneqq K_{h:\boldsymbol{b}}(\boldsymbol{\zeta}_1 - \boldsymbol{v})\, K_{j:\boldsymbol{b}}(\boldsymbol{\zeta}_2 - \boldsymbol{v})\, u_{hq:\boldsymbol{b}}(\boldsymbol{\zeta}_1)\, u_{jr:\boldsymbol{b}}(\boldsymbol{\zeta}_2)$, *where* $\boldsymbol{\zeta}_1$ *and* $\boldsymbol{\zeta}_2$ *either coincide completely (bivariate), have one common component (trivariate), or have no common components (tetravariate). Let* $\kappa$ *be the number of variates, and let* $\mathrm{d}\boldsymbol{\zeta}(\kappa)$ *represent the corresponding* $\kappa$*-variate differential. Then,*

$$\int_{\mathbb{R}^\kappa} (b_1 b_2)\, \mathcal{K}_{qr,hj:\boldsymbol{b}}(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2)\, \mathrm{d}\boldsymbol{\zeta}(\kappa) = \begin{cases} u_{hq:\boldsymbol{b}}(\boldsymbol{v})\, u_{jr:\boldsymbol{b}}(\boldsymbol{v}) \int_{\mathbb{R}^2} K(\boldsymbol{w})^2\, \mathrm{d}\boldsymbol{w} + O(b_1 \vee b_2) & \kappa = 2, \\ O(b_1 \wedge b_2) & \kappa = 3, \\ O(b_1 b_2) & \kappa = 4. \end{cases}$$

*Proof.* Recalling the definition of $K_{h:\boldsymbol{b}}(\boldsymbol{y}_h - \boldsymbol{v})$ from eq. (B.28), the integral in item (a) can be written as

$$\int_{\mathbb{R}^2} \sqrt{b_1 b_2} \cdot \frac{1}{b_1 b_2} K\left(\frac{\zeta_1 - v_1}{b_1}, \frac{\zeta_2 - v_2}{b_2}\right) u_{hq:\boldsymbol{b}}(\zeta_1, \zeta_2)\, \mathrm{d}\zeta_1\, \mathrm{d}\zeta_2, \tag{C.20}$$

which implies that the substitutions $w_1 = (\zeta_1 - v_1)/b_1$ and $w_2 = (\zeta_1 - v_2)/b_2$ gives the integral

$$\begin{aligned} &\int_{\mathbb{R}^2} \frac{\sqrt{b_1 b_2}}{b_1 b_2} K(w_1, w_2)\, u_{hq:\boldsymbol{b}}(b_1 w_1 + v_1, b_2 w_2 + v_2)\, (b_1 \mathrm{d}w_1)\, (b_2 \mathrm{d}w_2) \\ &= \sqrt{b_1 b_2} \cdot \int_{\mathbb{R}^2} K(w_1, w_2)\, u_{hq:\boldsymbol{b}}(b_1 w_1 + v_1, b_2 w_2 + v_2)\, \mathrm{d}w_1 \mathrm{d}w_2. \end{aligned} \tag{C.21}$$

Since $u_{hq:\boldsymbol{b}}(\boldsymbol{w})$ is a bivariate polynomial, it is clear that $u_{hq:\boldsymbol{b}}(b_1 w_1 + v_1, b_2 w_2 + v_2)$ can be written as

$$u_{hq:\boldsymbol{b}}(v_1, v_2) + b_1 c_1 w_1 + b_2 c_2 w_2 + b_1^2 c_{11} w_1^2 + b_1 b_2 c_{12} w_1 w_2 + b_2^2 c_{22} w_2^2, \tag{C.22}$$

for suitable constants $c_1$, $c_2$, $c_{11}$, $c_{12}$ and $c_{22}$. The integral in eq. (C.21) can thus be expressed as a sum of integrals like those occurring in lemma C.5(a), all of which are finite. The dominant term becomes $O\big(\sqrt{b_1 b_2}\big)$ when $\boldsymbol{b} \to \boldsymbol{0}^+$, and the conclusion of item (a) follows.

The substitution used in item (a) can also be applied for item (b), resulting in

$$\begin{aligned} &\left(\int_{\mathbb{R}^2} \left|\sqrt{b_1 b_2} \cdot \frac{1}{b_1 b_2} K(w_1, w_2)\, u_{hq:\boldsymbol{b}}(b_1 w_1 + v_1, b_2 w_2 + v_2)\right|^{\nu} (b_1 \mathrm{d}w_1)\, (b_2 \mathrm{d}w_2)\right)^{1/\nu} \\ &= |b_1 b_2|^{(2-\nu)/2\nu} \left(\int_{\mathbb{R}^2} \left|K(w_1, w_2)\, u_{hq:\boldsymbol{b}}(b_1 w_1 + v_1, b_2 w_2 + v_2)\right|^{\nu} \mathrm{d}w_1 \mathrm{d}w_2\right)^{1/\nu}. \end{aligned} \tag{C.23}$$

Note that this represent the norm in $L^{\nu}$-space, and that eq. (C.22) implies that it can be realised as the norm of a sum of the simpler components encountered in lemma C.5(c). It is now clear that Minkowski's inequality can be used to obtain a bound for the expression in eq. (C.23). In particular, constants $e_1$, $e_2$, $e_{11}$, $e_{12}$ and $e_{22}$ can be found that realises this bound as

$$|b_1 b_2|^{(2-\nu)/2\nu} \left( u_{hq:\boldsymbol{b}}(v_1, v_2) + b_1 e_1 w_1 + b_2 e_2 w_2 + b_1^2 e_{11} w_1^2 + b_1 b_2 e_{12} w_1 w_2 + b_2^2 e_{22} w_2^2 \right), \tag{C.24}$$

which is dominated by the $|b_1 b_2|^{(2-\nu)/2\nu}$-term when $\boldsymbol{b} \to \mathbf{0}^+$, as stated in item (b).

The investigation of item (c) requires different substitutions depending on the $\kappa$ for the configuration under investigation. Noting that the integrand in addition to the scaling factor $b_1 b_2$ always contains the product $K_{h:\boldsymbol{b}}(\boldsymbol{\zeta}_1 - \boldsymbol{v})\, K_{j:\boldsymbol{b}}(\boldsymbol{\zeta}_2 - \boldsymbol{v})$, it follows that it regardless of the value of $\kappa$ will be a factor $1/b_1 b_2$ that will be adjusted by the $b_1$- and $b_2$-factors that originates from the substituted differentials. It is easy to check that the new differentials becomes $b_1 b_2 \, \mathrm{d}w_1 \mathrm{d}w_2$ when $\kappa = 2$, $b_1^2 b_2 \, \mathrm{d}w_1 \mathrm{d}w_2 \mathrm{d}w_3$ or $b_1 b_2^2 \, \mathrm{d}w_1 \mathrm{d}w_2 \mathrm{d}w_3$ when $\kappa = 3$, and $b_1^2 b_2^2 \, \mathrm{d}w_1 \mathrm{d}w_2 \mathrm{d}w_3 \mathrm{d}w_4$ when $\kappa = 4$.

For the bivariate case, the substitution from item (a) gives an expression of the following form,

$$\int_{\mathbb{R}^2} K(w_1, w_2)^2 \cdot \mathcal{U}(w_1, w_2) \, \mathrm{d}w_1 \mathrm{d}w_2, \tag{C.25}$$

where $\mathcal{U}(w_1, w_2)$ is a product whose factors both are of the form encountered in eq. (C.22), i.e. it will be a quartic polynomial in the variables $(b_1 w_1)$ and $(b_2 w_2)$, and its constant term will be $u_{hq:\boldsymbol{b}}(\boldsymbol{v})\, u_{jr:\boldsymbol{b}}(\boldsymbol{v})$. From lemma C.6(b) it follows that this will be a finite integral, and as $\boldsymbol{b} \to \mathbf{0}^+$ the result will be as given for the $\kappa = 2$ case of item (c).

For the trivariate case, the overlap between $\boldsymbol{\zeta}_1$ and $\boldsymbol{\zeta}_2$ will belong to one of the following configurations, (i) $\boldsymbol{\zeta}_1 = (\zeta_1, \zeta_2)$ and $\boldsymbol{\zeta}_2 = (\zeta_1, \zeta_3)$, (ii) $\boldsymbol{\zeta}_1 = (\zeta_1, \zeta_2)$ and $\boldsymbol{\zeta}_2 = (\zeta_3, \zeta_1)$, (iii) $\boldsymbol{\zeta}_1 = (\zeta_1, \zeta_2)$ and $\boldsymbol{\zeta}_2 = (\zeta_2, \zeta_3)$, or (iv) $\boldsymbol{\zeta}_1 = (\zeta_1, \zeta_2)$ and $\boldsymbol{\zeta}_2 = (\zeta_3, \zeta_2)$. The reasoning is identical for the four cases, so it is sufficient to consider case (i), which gives the following product of kernel functions in the original integral,

$$K((\zeta_1 - v_1)/b_1, (\zeta_2 - v_2)/b_2) \cdot K((\zeta_2 - v_1)/b_1, (\zeta_3 - v_2)/b_2). \tag{C.26}$$

When the substitution

$$w_1 = (\zeta_1 - v_1)/b_1, \qquad w_2 = (\zeta_2 - v_2)/b_2, \qquad w_3 = (\zeta_3 - v_2)/b_2, \tag{C.27}$$

is used, the following component occurs in the transformed integrand,

$$\mathcal{K}\left(w_1, w_2, w_3\right) \coloneqq K(w_1, w_2) \cdot K(\left[(b_2 w_2 + v_2) - v_1\right] / b_1, w_3)\,. \tag{C.28}$$

The argument $\left[(b_2 w_2 + v_2) - v_1\right] / b_1$ does not pose a problem due to the boundedness requirement from eq. (B.27d) in definition B.9, and the following inequality thus holds for $\ell \in \{0, 1, 2\}$,

$$\begin{aligned}
\int_{\mathbb{R}^1} \mathcal{K}\left(w_1, w_2, w_3\right) w_3^\ell \,\mathrm{d}w_3 &= K(w_1, w_2) \cdot \int_{\mathbb{R}^1} K(\left[(b_2 w_2 + v_2) - v_1\right] / b_1, w_3)\, w_3^\ell \,\mathrm{d}w_3 && \text{(C.29a)}\\
&= K(w_1, w_2) \cdot \mathcal{K}_{2:\ell}(\left[(b_2 w_2 + v_2) - v_1\right] / b_1) && \text{(C.29b)}\\
&\leq \mathcal{D}_{2:\ell} \cdot K(w_1, w_2)\,, && \text{(C.29c)}
\end{aligned}$$

where $\mathcal{D}_{2:\ell}$ is a constant that bounds the function $\mathcal{K}_{2:\ell}$.

Since the substitution in eq. (C.27) transforms the integral of interest into

$$b_2 \int_{\mathbb{R}^3} \mathcal{K}\left(w_1, w_2, w_3\right) \cdot \mathcal{U}(w_1, w_2, w_3) \,\mathrm{d}w_1 \mathrm{d}w_2 \mathrm{d}w_3, \tag{C.30}$$

where $\mathcal{U}(w_1, w_2, w_3)$ is a quadratic polynomial in the variables $(b_1 w_1)$ and $(b_2 w_3)$, and a quartic polynomial in $w_2$ (with coefficients having suitable powers of $b_1$ and $b_2$ as factors), the observation in eq. (C.29) implies that an iterated approach to the integral (starting with the $w_3$-variable) can be used to show that each part of the sum will be bounded by a constant times an integral of the form encountered in lemma C.6(a). The trivariate integral in item (c) can thus be bounded by a sum of finite integrals having coefficients based on powers of $b_1$ and $b_2$. From the $b_2$ factor in eq. (C.30), it follows that the trivariate integral in this case is $O(b_2)$ when $\boldsymbol{b} \to \mathbf{0}^+$. Note that $w_2 = (\zeta_2 - v_1)/b_1$ could have been used as an alternative substitution in eq. (C.27), which by the obvious modifications of the arguments implies that the integral also will be $O(b_1)$ when $\boldsymbol{b} \to \mathbf{0}^+$ — and from this if follows that the integral is $O(b_1 \wedge b_2)$, which completes the proof for the $\kappa = 3$ case of item (c).

The case $\kappa = 4$ is quite simple, since no common components in $\boldsymbol{\zeta}_1$ and $\boldsymbol{\zeta}_2$ implies that the tetravariate integral, after the obvious substitution, corresponds to an expression of the form

$$b_1 b_2 \left( \int_{\mathbb{R}^2} K(\boldsymbol{w}) u_{hq:\boldsymbol{b}}(\boldsymbol{\zeta}(\boldsymbol{w})) \,\mathrm{d}\boldsymbol{w} \right) \cdot \left( \int_{\mathbb{R}^2} K(\boldsymbol{w}) u_{jr:\boldsymbol{b}}(\boldsymbol{\zeta}(\boldsymbol{w})) \,\mathrm{d}\boldsymbol{w} \right), \tag{C.31}$$

where $\boldsymbol{\zeta}(\boldsymbol{w}) = (b_1 w_1 + v_1, b_2 w_2 + v_2)$. The integrals occurring in this product are similar to those encountered in the bivariate case discussed above, and it is clear that the result will be $O(b_1 b_2)$ when $\boldsymbol{b} \to \mathbf{0}^+$, which concludes the proof of item (c). □

Note that the bivariate case of lemma C.6(c) only considers the configuration where the components of $\boldsymbol{\zeta}_1$ and $\boldsymbol{\zeta}_2$ coincide completely, while the configuration where $\boldsymbol{\zeta}_1 = (\zeta_1, \zeta_2)$ and $\boldsymbol{\zeta}_2$ is the diagonal reflection $(\zeta_2, \zeta_1)$ has been left out. This restriction does not pose a problem for the asymptotic investigation of $\widehat{f}_{\boldsymbol{v}}^{\,m}(\omega)$ when the point $\boldsymbol{v} = (v_1, v_2)$ lies upon the diagonal, i.e. when $v_1 = v_2$, since the diagonal folding property ensures that it is sufficient to consider positive lags for the point $\boldsymbol{v}$ in this case. For the general case, where $v_1 \neq v_2$, the following adjusted version of lemma C.6(c) is needed, where one of the kernels use $\boldsymbol{v}$ and the other use the diagonally reflected point $\breve{\boldsymbol{v}} = (v_2, v_1)$.

**Lemma C.7.** *The following holds for $u_{hq:\boldsymbol{b}}(\boldsymbol{w})$ and $K_{h:\boldsymbol{b}}(\boldsymbol{y}_h - \boldsymbol{v})$ from definitions B.7 and B.9, when the point $\boldsymbol{v} = (v_1, v_2)$ does not coincide with its diagonal reflection $\breve{\boldsymbol{v}} = (v_2, v_1)$, i.e. $v_1 \neq v_2$.*
*Let $\mathcal{K}_{qr,hj:\boldsymbol{b}}(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2; \boldsymbol{v}, \breve{\boldsymbol{v}}) \coloneqq K_{h:\boldsymbol{b}}(\boldsymbol{\zeta}_1 - \boldsymbol{v})\, K_{j:\boldsymbol{b}}(\boldsymbol{\zeta}_2 - \breve{\boldsymbol{v}})\, u_{hq:\boldsymbol{b}}(\boldsymbol{\zeta}_1)\, u_{jr:\boldsymbol{b}}(\boldsymbol{\zeta}_2)$, where $\boldsymbol{\zeta}_1$ and $\boldsymbol{\zeta}_2$ either are diagonal reflections of each other (bivariate), have one common component (trivariate), or have no common components (tetravariate). Let $\kappa$ be the number of variates, and let $\mathrm{d}\boldsymbol{\zeta}(\kappa)$ represent the corresponding $\kappa$-variate differential. Then,*

$$\int_{\mathbb{R}^\kappa} (b_1 b_2)\, \mathcal{K}_{qr,hj:\boldsymbol{b}}(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2; \boldsymbol{v}, \breve{\boldsymbol{v}})\, \mathrm{d}\boldsymbol{\zeta}(\kappa) = \begin{cases} o(1) & \kappa = 2, \\ O(b_1 \wedge b_2) & \kappa = 3, \\ O(b_1 b_2) & \kappa = 4. \end{cases}$$

*Proof.* The statements for the trivariate and tetravariate cases are identical to those in lemma C.6(c), and so are the proofs, i.e. the same substitutions can be applied for the present cases of interest.

For the bivariate case, the substitution $w_1 = (\zeta_1 - v_1)/b_1$ and $w_2 = (\zeta_1 - v_2)/b_2$ gives that the integral $\int_{\mathbb{R}^2} K(w_1, w_2)^2 \cdot \mathcal{U}(w_1, w_2)\, \mathrm{d}w_1\, \mathrm{d}w_2$ from eq. (C.25) is replaced with a sum of integrals of the form,

$$\int_{\mathbb{R}^2} K(w_1 + (v_1 - v_2)/b_1, w_2 + (v_2 - v_1)/b_2) \cdot K(w_1, w_2) w_1^k w_2^\ell\, \mathrm{d}w_1\, \mathrm{d}w_2, \tag{C.32}$$

where $k, \ell \geq 0$ and $k + \ell \leq 4$. and the integrands of these integrals goes to zero when $\boldsymbol{b} \to \boldsymbol{0}^+$, due to the assumption that $v_1 \neq v_2$. To clarify: For a kernel function $K$ whose nonzero values occurs on a bounded region of $\mathbb{R}^2$, the integrand of eq. (C.32) will become identical to zero when $(v_1 - v_2)/b_1$ and $(v_2 - v_1)/b_2$ are large enough to ensure that at least one of the factors in the integrand must be zero. For the general case, first observe that the factors $K(w_1, w_2) w_1^k w_2^\ell$ are the integrands that occurs in lemma C.5(a), and the finiteness

of those integrals implies that these factors must go to zero at a sufficiently high rate when $w_1$ and $w_2$ are far from origo. The rate at which the individual kernel $K(w_1, w_2)$ goes to zero will of course be faster than that of the product $K(w_1, w_2)w_1^k w_2^\ell$, and together this implies that the integrand in eq. (C.32) must go to zero when $\boldsymbol{b} \to \mathbf{0}^+$, and the integral thus becomes asymptotically negligible. □

It is a straightforward (albeit somewhat tedious) exercise to verify that eq. (C.32) goes towards zero at an exponential rate when the kernel function $K(\boldsymbol{w})$ is the product normal kernel. The observation that the bivariate case of lemma C.7 is $o(1)$ can also be derived from the realisation that $K_{h:\boldsymbol{b}}(\boldsymbol{\zeta}_1 - \boldsymbol{v})$ and $K_{j:\boldsymbol{b}}(\boldsymbol{\zeta}_2 - \breve{\boldsymbol{v}})$ are entities that converge towards two different bivariate Dirac delta functions, and the limit of the integral becomes zero since these delta functions sifts out different points.

## C.5 A few details related to $\sigma$-algebras, $\alpha$-mixing and $L^\nu$-spaces

The following general definitions and basic observations are needed when e.g. results from Davydov [1968] and Volkonskii and Rozanov [1959] are used.

**Related $\sigma$-algebras**

The $\sigma$-algebras related to the process $\{Y_t\}_{t\in\mathbb{Z}}$, will be denoted

$$\mathcal{F}_t^s \coloneqq \sigma(Y_t, \ldots, Y_s)\,, \tag{C.33}$$

where $t$ and $s$ are allowed to take the values $-\infty$ and $+\infty$ respectively.

Note in particular, that if a new random variable is defined by means of a measurable function $\xi(\boldsymbol{y}_{\bar{m}})$ from $\mathbb{R}^{m+1}$ to $\mathbb{R}$, i.e. $\mathcal{Y}_{m:t} \coloneqq \xi(\boldsymbol{Y}_{\bar{m}:t})$, then $\mathcal{Y}_{m:t} \in \mathcal{F}_t^{t+m}$.

**Inheritance of $\alpha$-mixing**

The coefficients in the strong mixing property mentioned in assumption 2.1(b), is given by

$$\alpha(s\,|\,Y_t) \coloneqq \sup\left\{|\mathrm{P}(A\cap B) - \mathrm{P}(A)\,\mathrm{P}(B)| \,:\, -\infty < t < \infty,\ A \in \mathcal{F}_{-\infty}^t,\ B \in \mathcal{F}_{t+s}^\infty\right\}, \tag{C.34}$$

from which it is an easy task to verify that a derived process, like the $\mathcal{Y}_{m:t}$ mentioned above, will have an inherited $\alpha$-mixing coefficient that satisfies

$$\alpha(s\,|\,\mathcal{Y}_{m:t}) \leq \alpha(s-m\,|\,Y_t)\,. \tag{C.35}$$

This implies that the finiteness requirement in eq. (2.21) will be inherited by the process $\mathcal{Y}_{m:t}$, i.e. with $\nu$ and $a$ as introduced in assumption 2.1(b), the following holds true

$$\sum_{j=1}^{\infty} j^a \left[\alpha(j\,|\,\mathcal{Y}_{m:t})\right]^{1-2/\nu} < \infty. \tag{C.36}$$

**Related $L^\nu$-spaces**

Some inequalities are needed in the main proofs, and these inequalities can be verified by means of the simple connection between expectations and $L^\nu$-spaces outlined below.[11]

First of all, when a measure space $(\Omega, \mathcal{G}, \mu)$ is given, then for $1 \leq \nu < \infty$, the space $L^\nu \coloneqq L^\nu(\Omega, \mathcal{G}, \mu)$ is defined to be the class of measurable real functions $\zeta$ for which $|\zeta|^\nu$ is integrable, that is,

$$\zeta(z) \in L^\nu \quad \overset{\text{def}}{\iff} \quad \int_\Omega |\zeta(z)|^\nu \, \mathrm{d}\mu < \infty. \tag{C.37}$$

The $L^\nu$-spaces related to the processes $\boldsymbol{Y}_{h:t}$ and $\boldsymbol{Y}_{\bar{m}:t}$ will henceforth be denoted by

$$L_h^\nu \quad \text{— the } L^\nu \text{ spaces related to the densities } g_h, \tag{C.38a}$$
$$L_{\bar{m}}^\nu \quad \text{— the } L^\nu \text{ space related to the density } g_{\bar{m}}. \tag{C.38b}$$

These $L^\nu$ spaces are in fact Banach spaces, see e.g. Billingsley [2012, Section 19] for details, which means that they are complete normed vector spaces, with a $\nu$-norm defined by

$$\|\zeta(z)\|_\nu \coloneqq \left( \int_\Omega |\zeta(z)|^\nu \, \mathrm{d}\mu \right)^{1/\nu} = (\mathrm{E}[|\zeta(Z)|^\nu])^{1/\nu} \tag{C.39}$$

and the Minkowski's inequality (i.e. the triangle inequality for $L^\nu$-spaces) will play a central role in the investigation later on,

$$\|\zeta_1(z) + \zeta_2(z)\|_\nu \leq \|\zeta_1(z)\|_\nu + \|\zeta_2(z)\|_\nu . \tag{C.40}$$

The main reason for the introduction of these $L^\nu$-spaces are the following observation: With $Z$ a random variable on $(\Omega, \mathcal{G}, \mu)$, the definitions of expectation and $L^\nu$-spaces gives a sequence of equivalences

$$\mathrm{E}[|\zeta(Z)|^\nu] < \infty \quad \iff \quad \int_\Omega |\zeta(z)|^\nu \, \mathrm{d}\mu < \infty \quad \iff \quad \zeta(z) \in L^\nu. \tag{C.41}$$

**Lemma C.8.** *For a univariate time series $\{Y_t\}_{t\in\mathbb{Z}}$, with $\boldsymbol{Y}_{h:t}$ and $\boldsymbol{Y}_{\bar{m}:t}$ as defined in definition 2.7, and with $m$ bivariate functions $\zeta_h : \mathbb{R}^2 \longrightarrow \mathbb{R}^1$*

$$\begin{gathered} \textit{If } \mathrm{E}[|\zeta_h(\boldsymbol{Y}_{h:t})|^\nu] < \infty \textit{ for } h = 1, \ldots, m, \textit{ then} \\ \left(\mathrm{E}\left[\left|\textstyle\sum_{h=1}^m a_h \zeta_h(\boldsymbol{Y}_{h:t})\right|^\nu\right]\right)^{1/\nu} \leq \textstyle\sum_{h=1}^m |a_h| \left(\mathrm{E}[|\zeta_h(\boldsymbol{Y}_{h:t})|^\nu]\right)^{1/\nu} < \infty. \end{gathered}$$

[11] These definitions are normally presented with $p$ used instead of $\nu$.

*Proof.* From eq. (C.41) it follows that $\mathrm{E}[|\zeta_h(\boldsymbol{Y}_{h:t})|^\nu] < \infty$ implies $\zeta_h(\boldsymbol{y}_h) \in L_h^\nu$ for $h = 1, \ldots, m$. With $\widetilde{\zeta}_h(\boldsymbol{y}_{\bar{m}})$ the corresponding trivial extensions to $(m+1)$-variate functions, it follows from eq. (2.20) that $\widetilde{\zeta}_h(\boldsymbol{y}_{\bar{m}}) \in L_{\bar{m}}^\nu$ for $h = 1, \ldots, m$. From the vector space property of $L^\nu$-spaces it follows that $\sum_{h=1}^m a_h \zeta_h(\boldsymbol{Y}_{h:t}) \in L_{\bar{m}}^\nu$, and Minkowski's inequality then gives the desired result. □

# Appendix D: Sensitivity analysis of the tuning parameters

This section will investigate how sensitive $\widehat{f}_{\boldsymbol{v}}^m(\omega)$ is to changes in the tuning parameters (and the point $\boldsymbol{v}$). This will be done by the distance function $D$ introduced in appendix D.1, together with plots that reveal information about the frequency-dimension.

Appendices D.2 and D.3 respectively consider the sensitivity of the point $\boldsymbol{v}$ and the bandwidth $\boldsymbol{b}$, whereas the sensitivity of the truncation level $m$ is discussed in appendix D.4. The effect the value of the block length $L$ has upon the bootstrap-based pointwise confidence intervals is discussed in appendix F.5, since that gives the most natural flow.

Appendix E contains a discussion related to the *selection of tuning parameters* for $\widehat{f}_{\boldsymbol{v}}^m(\omega)$, and it also contains some references to the related problem of selecting the bandwidth when a local Gaussian correlation is to be estimated from a sample.

The scripts required for the replication of the results in this section are contained in the `R`-package `localgaussSpec`, and these scripts can be used as templates for those that would like to investigate other time series in a similar manner. See appendix G for details.

## D.1 Sensitivity analysis - the distance function

An investigation of the sensitivity requires a tool that can measure the differences that occur in the resulting estimates when the tuning parameters are adjusted. Many techniques have been developed in order to deal with distances between spectral functions, cf. e.g. Basseville [2013, Section 7] and Georgiou [2007, Section 1]. Some approaches are based on proper distance functions, whereas other use divergence/distortion measures where symmetry and the triangular identity no longer are present.

A natural (and easy to implement) candidate for the case of interest in this paper is the distance function inherited from the complex Hilbert space of Fourier series on the interval $\left[-\frac{1}{2}, \frac{1}{2}\right]$, cf. e.g. Brockwell and Davis [1986, Ch. 2.8], i.e. for $f(\omega) = \sum_{h=-\infty}^{\infty} \rho(h) e^{-2\pi i h}$ the norm is defined by $||f(\omega)||^2 = \int_{-1/2}^{1/2} f(\omega)\overline{f(\omega)}\,\mathrm{d}\omega = \sum_{h=-\infty}^{\infty} \rho(h)^2$. This motivates the following definition.

**Definition D.1.** *Given two spectra* $f_1(\omega) = \sum_{h=-\infty}^{\infty} \rho_1(h)e^{-2\pi ih}$ *and* $f_2(\omega) = \sum_{h=-\infty}^{\infty} \rho_2(h)e^{-2\pi ih}$, *the distance between them is denoted by*

$$D(f_1(\omega), f_2(\omega)) := \sqrt{\sum_{h=-\infty}^{\infty} (\rho_1(h) - \rho_2(h))^2}. \tag{D.1}$$

*Furthermore: The notation* $D(f_1(\omega))$ *will be interpreted as* $D(f_1(\omega), 0)$, *which implies that* $D(f_1(\omega), f_2(\omega))$ *also can be written as* $D(f_1(\omega) - f_2(\omega))$ *(which is used in fig. D.8).*

Note that $D$ will work both for real-valued and complex-valued spectra, and the latter is of importance both with regard to the univariate case when the point $\boldsymbol{v}$ lies of the diagonal, and with regard to the multivariate case treated in Jordanger and Tjøstheim [2017] .

The obvious adjustment must be done when $D$ is used on $m$-truncated estimates $\widehat{f}_{\boldsymbol{v}}^{\,m}(\omega)$, i.e. $\rho(h)$ should be replaced with $\lambda_m(h) \cdot \widehat{\rho}_{\boldsymbol{v}}(h)$ when $|h| \leq m$, and with 0 when $|h| > m$.

The distance function in definition D.1 is not applicable in the FFT-periodogram based approach to the estimation of spectral densities, since that approach does not explicitly compute the coefficients needed in eq. (D.1). However, note that the deviance measure that is based on the *root mean squared error* (RMSE), cf. e.g. Chen et al. [2019, Section 3.2], is closely related to the one used in the present paper. To emphasise: If $\tilde{f}_1(\omega)$ and $\tilde{f}_2(\omega)$ are two periodogram-based estimates of the spectral densities $f_1(\omega)$ and $f_2(\omega)$, then the RMSE-distance is given by

$$D_{\text{RMSE}}\left(\tilde{f}_1(\omega), \tilde{f}_2(\omega)\right) := \sqrt{\frac{1}{n}\sum_{l=0}^{n-1}\left[\tilde{f}_1(\omega_l) - \tilde{f}_2(\omega_l)\right]^2}, \tag{D.2}$$

where the summation is over all the Fourier-frequencies $\omega_l = l/n$ in the interval $[0, 1)$. A quick inspection of the expression under the square-root in eq. (D.2) reveals that this is a Riemann-sum approximation of the integral $\int_0^1 (f_1(\omega) - f_2(\omega))\,\overline{(f_1(\omega) - f_2(\omega))}\,\mathrm{d}\omega$. This will, when $n \to \infty$, converge towards $||f_1(\omega) - f_2(\omega)||^2$, which shows the close connection with the distance function from definition D.1.

Another more commonly used divergence measure is also considered in Chen et al. [2019], and that is the divergence measure based on the Kullback-Leibler (KL) divergence Kullback and Leibler [1951]. For the periodogram-based approach this can be written as

$$D_{\text{KL}}\left(\tilde{f}_1(\omega), \tilde{f}_2(\omega)\right) := \sum_{l=0}^{n-1} \tilde{f}_1(\omega_l) \ln\left(\frac{\tilde{f}_1(\omega_l)}{\tilde{f}_2(\omega_l)}\right). \tag{D.3}$$

An implementation of the KL-approach in this paper was briefly considered, but it was discarded since the local Gaussian spectral densities $f_v(\omega)$ in general will be complex-valued functions, and it was thus not clear how to adjust eq. (D.3) in a proper manner.

**Regarding the frequency-dimension:** A distance measure like the $D$ defined in definition D.1 does not contain any information about the frequencies, and completely different spectral densities can have the same distance-value. It is thus, for the purpose of sensitivity analysis, important to combine distance-based plots with plots that reveal something about the frequency-component too.

## D.2 Sensitivity analysis: The point $\boldsymbol{v}$

The bandwidth $\boldsymbol{b}$ is a central tuning parameter when an estimate of the $m$-truncated local Gaussian spectral density $f_{\boldsymbol{v}}^m(\omega)$ is desired for a given point $\boldsymbol{v} = (v_1, v_2)$. The point $\boldsymbol{v}$ itself is not a tuning parameter of the estimation algorithm, but an investigator will obviously be interested in information about how $f_{\boldsymbol{v}}^m(\omega)$ varies with $\boldsymbol{v}$, and it is thus also natural to consider the sensitivity of the estimate $\widehat{f}_{\boldsymbol{v}}^m(\omega)$ relatively the selected point.

Two plots related to this particular investigation was included in the main part, i.e. figs. 8 and 10, which respectively considered the *local trigonometrical example* and the `dmbp`-data. It is preferable to have a plot available for the present discussion too, and fig. D.1 shows an example based on one single simulation from the apARCH$(2, 3)$ that was fitted to the `dmbp`-data, cf. section 3.4.

The point $\boldsymbol{v} = (v_1, v_2)$ is bivariate, but the present investigation will restrict its attention to the diagonal cases. The requirement $v_1 = v_2$ is used for simplicity since it ensures that the resulting local Gaussian spectral densities will be real valued.

This restriction implies that the point $\boldsymbol{v}$ is allowed to vary continuously along a one dimensional line (the diagonal), and a heatmap can be used to see how $\widehat{f}_{\boldsymbol{v}}^m(\omega)$ varies with $\boldsymbol{v}$ (for a fixed $\boldsymbol{b}$). It is also of interest to use the distance function $D$ from definition D.1 to create a distance-based plot that shows how the norm $D\Big(\widehat{f}_{\boldsymbol{v}}^m(\omega)\Big)$ varies with $\boldsymbol{v}$.

The points $\boldsymbol{v}$ in fig. D.1 ranges from the 5% percentile to the 95% percentile of the standard normal distribution, increasing in steps of 0.5% (altogether 91 different points). This percentile based selection implies that the corresponding points are not equally spaced along the actual diagonal, and the plots in fig. D.1 have thus used the option that the points $\boldsymbol{v}$ have been presented according to their underlying percentile-values — which implies that these plots primarily reveals information about the copula-structure of the time series under investigation.

It can be seen from fig. D.1 that $\widehat{f}_{\boldsymbol{v}}^m(\omega)$ near the 50% percentile is quite close to an i.i.d. white noise situation — and it also seems to be a very clear symmetry around the

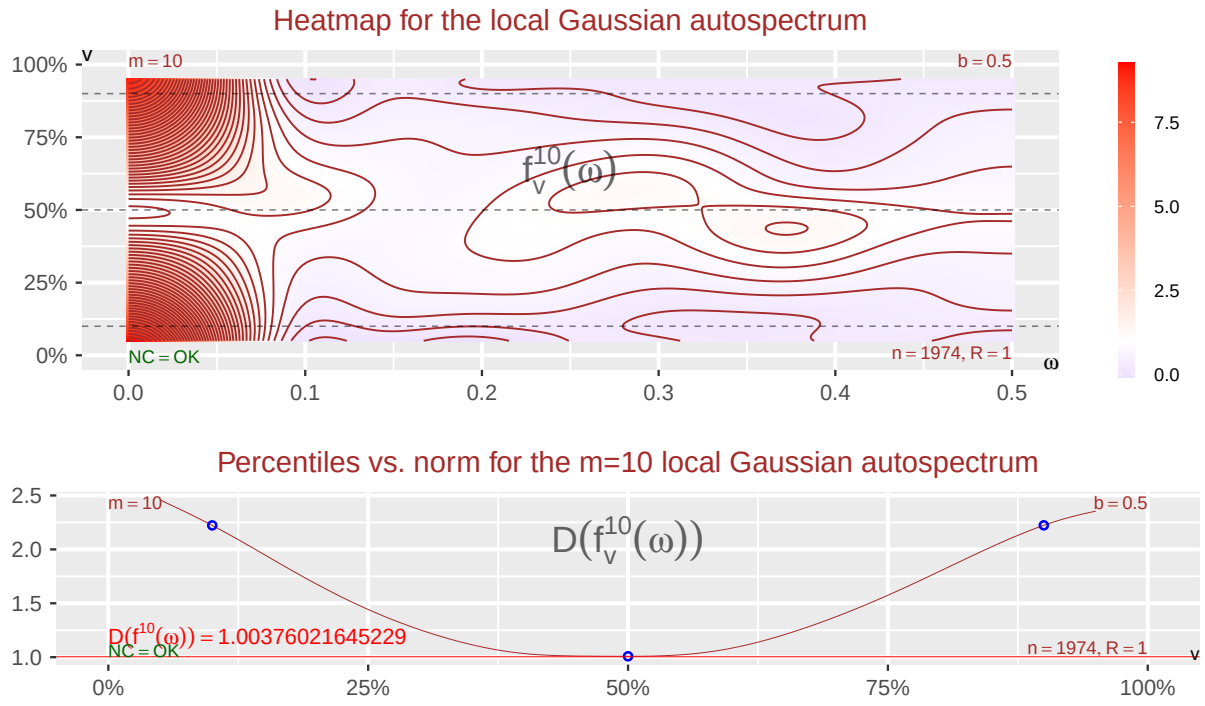


Figure D.1: Heatmap and corresponding distance-based plots for the the apARCH(2, 3)-model that was fitted to the `dmbp`-data, showing how $\widehat{f}_{\boldsymbol{v}}^{10}(\omega)$ varies with the percentiles for the diagonal-points $\boldsymbol{v}$. The percentiles used in fig. 11, i.e. 10%, 50% and 90%, have been highlighted with lines/points.

50% percentile. This is in stark contrast to the situation seen for the `dmbp`-data, cf. fig. 10, which indicates an asymmetry around the 50% percentile

Note that the 5% and 95% percentiles are quite far out in the tails of the distribution, and it is thus natural to assume that the selected bandwidth in those cases might fail to work properly — the small sample variation of the points closest to the point $\boldsymbol{v}$ might simply render the estimated local Gaussian autocorrelations rather dubious. It is possible to counter this problem by selecting a larger bandwidth for percentiles in the tails, but it is then important to keep in mind that a too large bandwidth might completely miss the desired local structure at the point of investigation.

**Heatmap-plots for the estimates $\widehat{\rho}_{\boldsymbol{v}}(h)$:** The construction of the two plots in fig. D.1 requires the computation of all of the underlying estimates $\widehat{\rho}_{\boldsymbol{v}}(h)$, for $h = 1, \ldots, m$. It is thus also possible to create heatmap-based plots that can visualise how these estimates changes as the point $\boldsymbol{v}$ moves from the 5% to the 95% percentile, and this can for the apARCH$(2, 3)$-example be seen in fig. D.2.

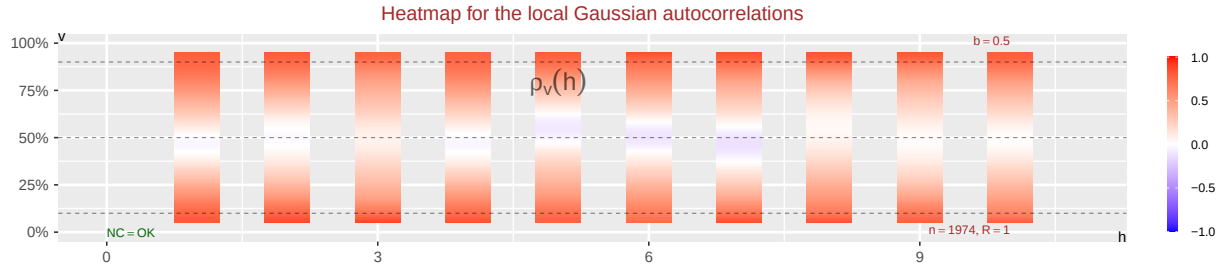


Figure D.2: Heatmap for $\widehat{\rho}_{\boldsymbol{v}}(h)$, for the apARCH$(2, 3)$-model fitted to the `dmbp`-data. The percentiles used in fig. 11, i.e. 10%, 50% and 90%, have been highlighted with lines.

It is clear from fig. D.2 that the estimated values $\widehat{\rho}_{\boldsymbol{v}}(h)$ are near symmetric around the 50% percentile, which thus explains the corresponding symmetry for $\widehat{f}^{\,m}_{\boldsymbol{v}}(\omega)$ seen in fig. D.1. For the `dmbp`-data, see fig. D.3, a similar level of symmetry is not to the same extent present. It might from such plots be possible to identify if it is the contribution from some particular lags $h$ that drives the asymmetry of the corresponding estimated spectrum $\widehat{f}^{\,m}_{\boldsymbol{v}}(\omega)$.

For completeness, fig. D.4 has been included in order to show how the situation looks like for the *local trigonometric example* seen in fig. 8.

## D.3 Sensitivity analysis: The bandwidth $b$

The bandwidth $\boldsymbol{b} = (b_1, b_2)$ is bivariate, but it is natural to assume $b_1 = b_2$ when a univariate time series is investigated. With this restriction it follows that the sensitivity of $\widehat{f}^{\,m}_{\boldsymbol{v}}(\omega)$ due to changes in the bandwidth $\boldsymbol{b}$ can be investigated in a similar manner to the one used in the preceding section for the diagonal points $\boldsymbol{v}$.

The bandwidth $\boldsymbol{b}$ should be selected according to the Goldilocks principle, i.e. it should neither be 'too low' nor 'too high', it must be 'just right'. The heatmap and distance-based plots from fig. D.1 can easily be adjusted to visualise the problems that occur when the bandwidth does not belong to the 'just right' region. The plots shown in fig. D.5 does

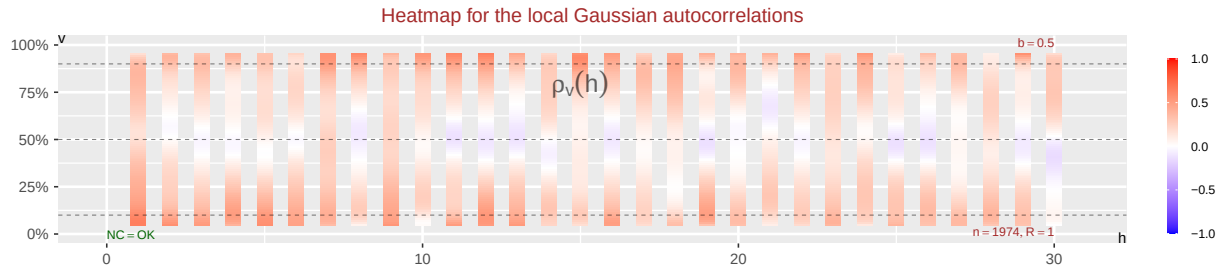


Figure D.3: Heatmap for $\widehat{\rho}_{\boldsymbol{v}}(h)$, for the `dmbp`-data. The percentiles used in fig. 9, i.e. 10%, 50% and 90%, have been highlighted with lines.

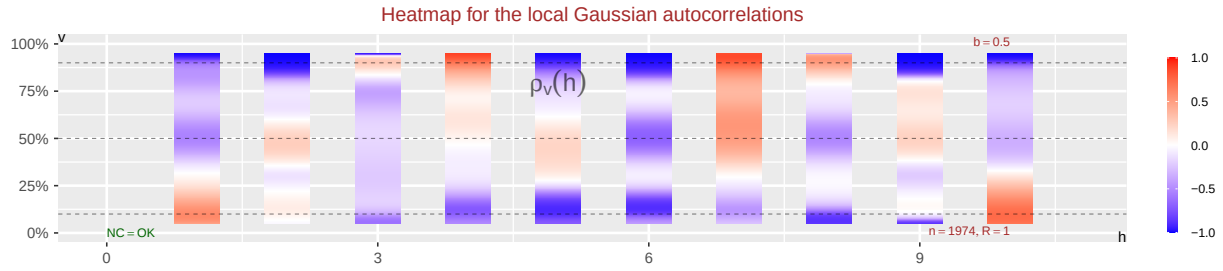


Figure D.4: Heatmap for $\widehat{\rho}_{\boldsymbol{v}}(h)$, for the *local trigonometric case*. The percentiles used in fig. 7, i.e. 10%, 50% and 90%, have been highlighted with lines.

once more consider the `dmbp`-data, and in this case the bandwidth ranges from 0.25 to 1.5 in steps by 0.005 (altogether 251 different bandwidths). The bandwidth $b = 0.5$ has been highlighted since it was that value that was used in fig. 9.

The problem when $\boldsymbol{b}$ becomes too large is that the estimated local Gaussian autocorrelations $\widehat{\rho}_{\boldsymbol{v}}(h)$ no longer will capture the local structure of interest, and the corresponding estimated local Gaussian spectral density $\widehat{f}^{m}_{\boldsymbol{v}}(\omega)$ (which no longer deserves to be referred to as 'local') will then be indistinguishable from the ordinary spectral density. It is clear from fig. D.5 that a bandwidth of $b = 1.5$ is far too large for the present investigation.

The expected behaviour when a too low bandwidth is used is that it will trigger a degeneration of the estimated local Gaussian autocorrelations, i.e. $\widehat{\rho}_{\boldsymbol{v}}(h)$ will tend towards either $+1$ or $-1$ regardless of the actual structure of the underlying density distributions.

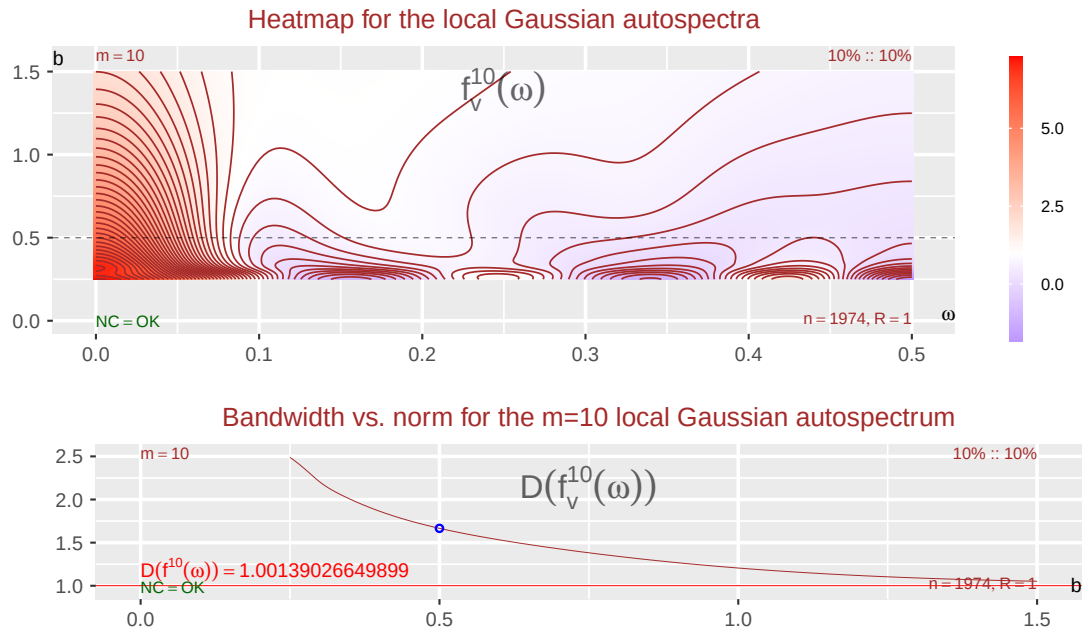


Figure D.5: Heatmap and corresponding distance-based plots based on the `dmbp-data`, showing how $\widehat{f}_{\boldsymbol{v}}^{10}(\omega)$ varies with the bandwidth $b$. The bandwidth used in fig. 9, i.e. $b = 0.5$, has been highlighted with a line/point.

The reason for this is that $\widehat{\rho}_{\boldsymbol{v}}(h)$ will, due to the kernel function from the density estimation algorithm, become increasingly sensitive to the position of the $h$-lagged pairs $\left(Y_{t+h}, Y_t\right)$ that lies nearest to the point $\boldsymbol{v} = (v_1, v_2)$. To clarify, for a given point $\boldsymbol{v}$ there will be a collection of Euclidean distances to the $h$-lagged pairs $\left(Y_{t+h}, Y_t\right)$ in the sample, and these distances could (after a re-indexing) be sorted in ascending order $\{d_i\}_{i=1}^{n-h}$.

Under the assumption that it is the product normal kernel that is used, the contribution from a lag-$h$ pair $\left(Y_{t+h}, Y_t\right)$ that lies a distance of $d_i$ from $\boldsymbol{v}$ will be weighted by $w_{i:\boldsymbol{b}} := \frac{1}{2\pi b^2} e^{-d_i^2/2b^2}$ — and it is now natural to consider the set of all the weights $\mathcal{W}_{\boldsymbol{v}:\boldsymbol{b}} := \{w_{i:\boldsymbol{b}}\}_{i=1}^{n-h}$.

The primary detail of interest is how much larger the weights are for the pairs that lies closest to $\boldsymbol{v}$, and it thus necessary to consider the fraction $r_{ij:\boldsymbol{b}} := w_{i:\boldsymbol{b}}/w_{j:\boldsymbol{b}} = \left(e^{d_j^2 - d_i^2}\right)^{1/b^2}$. The number $r_{ij:\boldsymbol{b}}$ will, when $d_i < d_j$, grow to $\infty$ when $\boldsymbol{b} \to \boldsymbol{0}^+$, and this implies that the estimation algorithm for small $b$-values will become increasingly sensitive to the $h$-lagged

pairs that lies closest to the point $\boldsymbol{v}$ when the bandwidth shrinks — and in the end it would thus be natural to have a degeneration of the estimated value $\rho_{\boldsymbol{v}}(h)$ to either $+1$ or $-1$.

Note that the corresponding $D\left(\widehat{f}_{\boldsymbol{v}}^{\,m}(\omega)\right)$ will grow when this degeneration happens, as can be seen for $b = 0.25$ in the distance-based plot in fig. D.5.

**Heatmap-plots for the estimates $\widehat{\rho}_{\boldsymbol{v}}(h)$:** It is here, as it was for the investigation of the diagonal points $\boldsymbol{v}$, possible to also consider a heatmap based investigation of the underlying estimates $\widehat{\rho}_{\boldsymbol{v}}(h)$, for $h = 1, \ldots, m$. Such a plot is given in fig. D.6, and it can there be observed that it for some of $\widehat{\rho}_{\boldsymbol{v}}(h)$-estimates is the case that the estimates first switch sign from positive to negative — and then they grows quickly towards $-1$. This kind of behaviour is expected to occur when the bandwidth $\boldsymbol{b}$ has shrunk to a level that implies that the kernel function in the local penalty function, cf. eq. (B.7), gives high weights to the few observations $\left(Y_{t+h}, Y_t\right)$ nearest $\boldsymbol{v}$, and very low weights elsewhere.

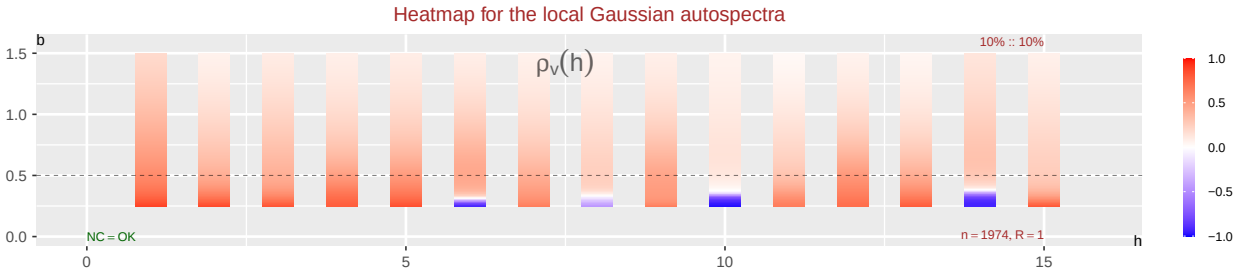


Figure D.6: Heatmap for $\widehat{\rho}_{\boldsymbol{v}}(h)$, for the `dmbp`-data, showing the effect of different bandwidths. The bandwidth used in fig. 9, i.e. 0.5 has been highlighted with a line.

Note that fig. D.5 considers the situation where $\boldsymbol{v}$ is the diagonal point corresponding to the lower tail, but similar plots could have been included for the cases where $\boldsymbol{v}$ corresponds to either the center or the upper tail.[12] A comparison of the distance-based plots for these three points is presented in fig. D.7, in order to show how the bandwidth-sensitivity of $\widehat{f}_{\boldsymbol{v}}^{\,m}(\omega)$ also depends on the selected point $\boldsymbol{v}$. A common scale has been used for the three subplots in order to emphasise the asymmetry between the lower and upper tail.

Note that the center plot of fig. D.7 reveals that the 'too low bandwidth problem' occurs a bit slower in a high density region, but it will even there eventually create a situation where the estimated local Gaussian autocorrelations degenerate towards either $+1$ or $-1$.

[12] The interested reader can use the scripts in the `R`-package `localgaussSpec` to get access to these plots for the center and upper tail, cf. appendix G for details.

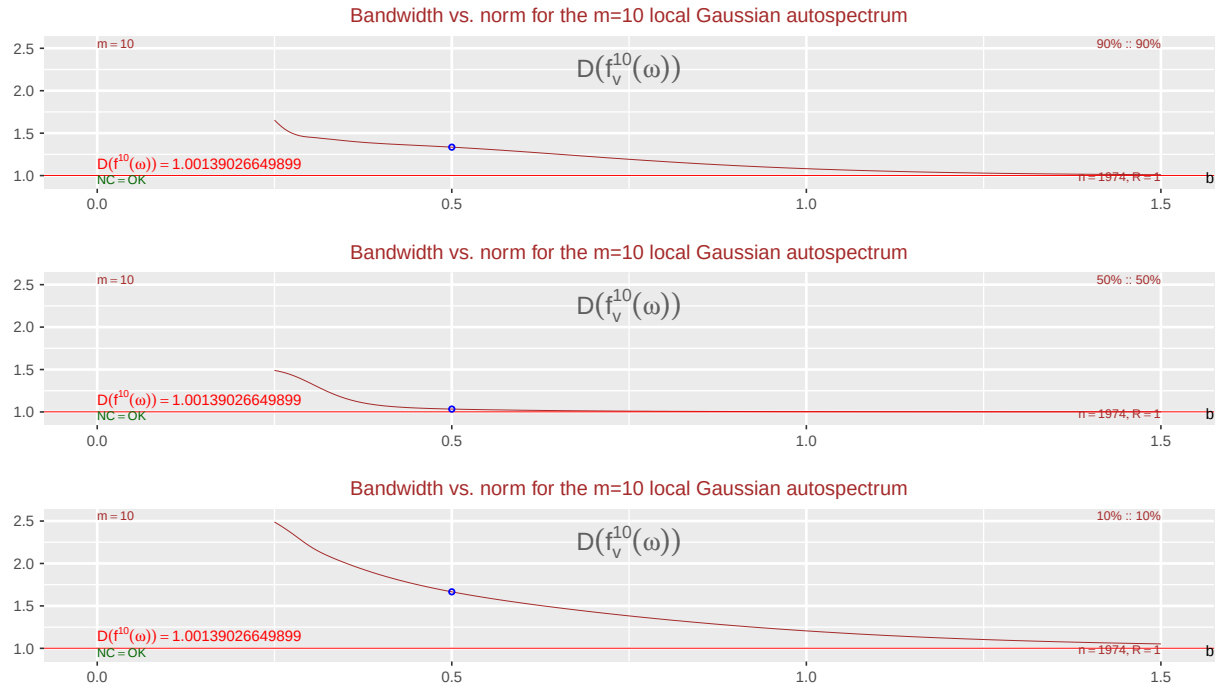


Figure D.7: Distance-based plots for the `dmbp-data`, showing how $\widehat{f}_{\boldsymbol{v}}^{10}(\omega)$ varies with the bandwidth $b$ for the three percentiles used in fig. 9, i.e. 10%, 50% and 90%. The bandwidth used in fig. 9, i.e. $b = 0.5$, has been highlighted by a point.

The heatmap and distance-based plots in figs. D.5 to D.7 can detect the clearly undesirable regions for the bandwidth $\boldsymbol{b}$, but they do not reveal what the 'just right' value for the bandwidth should be. Nevertheless, it is still possible to gain some insight into how sensitive the estimate of $f_{\boldsymbol{v}}^{m}(\omega)$ will be for minor variations of the bandwidth $\boldsymbol{b}$, and that can be useful with regard to the selection of a few bandwidths that can be used when e.g. a bootstrap-investigation is to be performed.

The framework used in the `R`-package `localgaussSpec` ensures that it is trivial to compute and investigate a wide range of bandwidths simultaneously, and the key idea is that knowledge of the local dependency structure can still be obtained even if the selected bandwidths are not spot on the 'just right' value for the bandwidth.

## D.4 Sensitivity analysis: The truncation level $m$

The shape of $f_{\boldsymbol{v}}^{m}(\omega)$ for a low truncation level can be different from the shape seen when a higher truncation level is used. It is thus of interest to investigate how sensitive the estimates $\widehat{f}_{\boldsymbol{v}}^{m}(\omega)$ are to changes in the truncation level $m$.

This issue can easily be probed by performing an initial investigation with a high value for the maximum lag to be computed, since the computational cost is not too large when only a single sample (like the `dmbp`-data) is investigated. It did e.g. not take a long time to estimate $\rho_{\boldsymbol{v}}(h)$ for $h = 1, \ldots, 200$, which was needed for the construction of fig. 4 in the main document — and with these estimates it is trivial to compare $\widehat{f}^{m}(\omega)$ and $\widehat{f}_{\boldsymbol{v}}^{m}(\omega)$ for $m$ up to 200, since the integrated `shiny`-application of the `R`-package `localgaussSpec` can animate the changes that occur in the spectra when $m$ grows from 0 to 200.

The computational costs can become rather large when it is necessary to find pointwise confidence intervals, since a high number of replicates then must be investigated with the same configuration of tuning parameters. It is then important to figure out a sufficient truncation level $m$, and restrict the attention to the estimates of $\rho_{\boldsymbol{v}}(h)$ for $h = 1, \ldots, m$.

A drawback with the `shiny`-based approach in `localgaussSpec` is that it requires an inspection of many different plots. It could thus be of interest to also consider summary-plots that either use the distance function $D$ from definition D.1, or some heatmap-based alternative visualisation of $\widehat{f}_{\boldsymbol{v}}^{m}(\omega)$, similar to those used for $\widehat{\rho}_{\boldsymbol{v}}(h)$ in figs. D.2 to D.4 and D.6.

**Distance plots:** It is possible to investigate the $m$-sensitivity by distance-based plots, but those plots are less useful in this case. One reason for this is that the norms $D\Big(\widehat{f}_{\boldsymbol{v}}^{m}(\omega)\Big)$ are monotonically increasing as functions of $m$. This can easily be seen by first recalling (cf. algorithm 2.5(c)) that the estimates $\widehat{f}_{\boldsymbol{v}}^{m}(\omega)$ are given by

$$\widehat{f}_{\boldsymbol{v}}^{m}(\omega) \coloneqq 1 + \sum_{h=1}^{m} \lambda_{m}(h) \cdot \widehat{\rho}_{\breve{\boldsymbol{v}}}(h) \cdot e^{+2\pi i \omega h} + \sum_{h=1}^{m} \lambda_{m}(h) \cdot \widehat{\rho}_{\boldsymbol{v}}(h) \cdot e^{-2\pi i \omega h},$$

and then keeping in mind that the lag-window function $\lambda_{m}(h)$ satisfies $\lambda_{m+1}(h) \geq \lambda_{m}(h)$. It follows that $D\Big(\widehat{f}_{\boldsymbol{v}}^{m+1}(\omega)\Big) \geq D\Big(\widehat{f}_{\boldsymbol{v}}^{m}(\omega)\Big)$, which does not provide any useful new information.

Instead of a plot of the norms $D\Big(\widehat{f}_{\boldsymbol{v}}^{m}(\omega)\Big)$, it is slightly more interesting to consider a plot that shows $D\Big(\widehat{f}_{\boldsymbol{v}}^{m+1}(\omega) - \widehat{f}_{\boldsymbol{v}}^{m}(\omega)\Big)$, i.e. the distances between $\widehat{f}_{\boldsymbol{v}}^{m+1}(\omega)$ and $\widehat{f}_{\boldsymbol{v}}^{m}(\omega)$ in the Hilbert room of Fourier series. This idea is shown in fig. D.8 for the three diagonal points and 200 lags that was included in fig. 4. Note that fig. D.8 takes into account the scaling due to the lag-window function $\lambda_{m}(h)$, and as such it does provide some new

information compared to that contained in the plot showing the estimated local Gaussian autocorrelations.

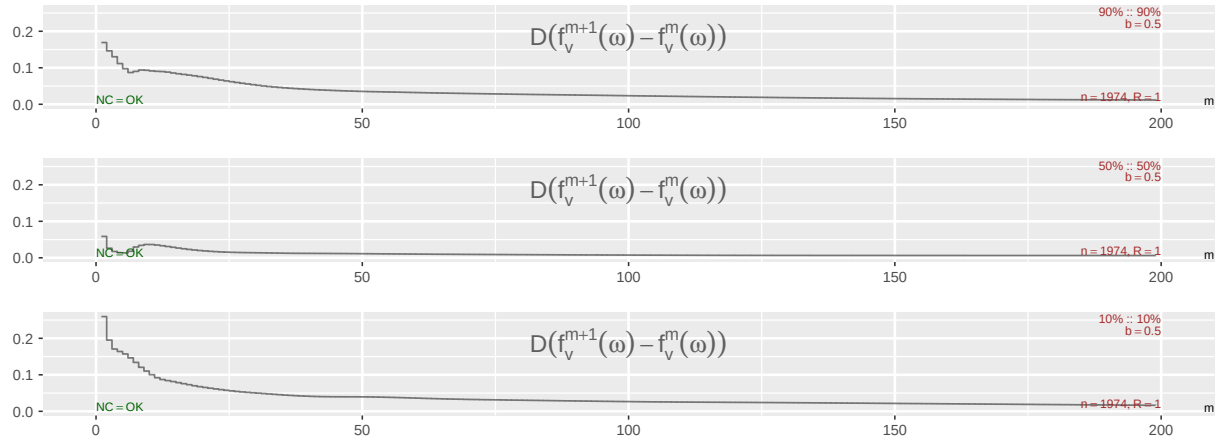


Figure D.8: Distances between successive $m$-truncations of the local spectra, `dmbp`-data.

The three subplots of fig. D.8 shows that $D\left(\widehat{f}_{\boldsymbol{v}}^{m+1}(\omega) - \widehat{f}_{\boldsymbol{v}}^{m}(\omega)\right)$ rather quickly starts to decrease monotonically, which is as expected given the presence of the lag-window function $\lambda_m(h)$. This decrease implies that the effect of a change in the truncation level from $m$ to $m+1$ becomes smaller as $m$ grows, and the sensitivity is thus largest when $m$ is small.

Figure D.8 might indicate that the $m = 10$ used in the main part is a bit to small. However, the purpose of that particular truncation level was simply to show that even a low truncation level could be used to detect the presence of nonlinear dependency structures in the time series under investigation, i.e. structures not detected by the ordinary spectrum.

It is natural to assume that two *successive* local Gaussian spectra $\widehat{f}_{\boldsymbol{v}}^{m}(\omega)$ and $\widehat{f}_{\boldsymbol{v}}^{m+1}(\omega)$ should be similar in shape when $m$ has grown a bit, but this does not imply that the *accumulated* changes to $\widehat{f}_{\boldsymbol{v}}^{m}(\omega)$ are negligible. It is thus important to also inspect the frequency-dimension, and this can as mentioned above easily be done by the interactive `shiny`-application in the `localgaussSpec`-package.

**Heatmap plots:** The truncation level $m$ is a discrete tuning parameter, and an inspection based on a heatmap-based approach could thus follow the setup used for the estimated $\widehat{\rho}_{\boldsymbol{v}}(h)$-values seen in figs. D.2 to D.4 and D.6. The `R`-package `localgaussSpec` contains a script that can be used to create such a heatmap-based plot for $\widehat{f}_{\boldsymbol{v}}^{m}(\omega)$, with the frequencies $\omega$ along one axis and the truncation levels $m$ along the other.

The resulting heatmap-based plot clearly showed that the peak seen in fig. 9 at $\omega = 0$ (for $m = 10$ and a point either in the lower or upper tail) became even more dominating as $m$ increased, and the peak dominated to such an extent that the heatmap-based plot did not reveal anything about the other frequencies. This plot has thus not been included here, but the script is available in `localgaussSpec`, cf. appendix G for details.

# Appendix E: How to select the tuning parameters?

Several tuning parameters are required in order to compute the $m$-truncated estimate $\widehat{f}_{\boldsymbol{v}}^{m}(\omega)$ of the local Gaussian spectrum $f_{\boldsymbol{v}}(\omega)$, for a given point $\boldsymbol{v}$. In addition to the truncation level $m$, there is a bandwidth $\boldsymbol{b}$ (to be used when estimating the local Gaussian autocorrelations $\rho_{\boldsymbol{v}}(h)$, for $h \in \{1, \ldots, m\}$). There is also a lag-window function $\lambda_m(h)$ used for smoothing.

The sensitivity analysis in appendix D considered the effect of minor changes to the tuning parameters $\boldsymbol{b}$ and $m$, and it did also discuss the sensitivity of $\widehat{f}_{\boldsymbol{v}}^{m}(\omega)$ that is due to the position of the point $\boldsymbol{v}$ — which is of interest to know when a given sample/model is to be investigated.

The task of finding 'optimal tuning parameters' lies beyond the scope of this paper, and the focal point of interest in this section will be to give some advice with regard to how the R-package `localgaussSpec` can be used to investigate a given sample/model, cf. appendix E.1 for the details. A few comments related to the selection of the bandwidth $\boldsymbol{b}$ is given in appendix E.2, primarily in order to give some pointers to papers that have discussed bandwidth selection for the estimation of the local Gaussian correlation $\rho_{\boldsymbol{v}}$.

## E.1 Using the R-package `localgaussSpec`

The R-package `localgaussSpec` can compute $\widehat{f}_{\boldsymbol{v}}^{m}(\omega)$ for a wide range of tuning parameters, and for a huge selection of different points $\boldsymbol{v}$. The integrated `shiny`-application enables an easy interactive investigation of the resulting estimates, with an interface that makes it trivial to switch between visualisations of the estimated local Gaussian autocorrelations $\widehat{\rho}_{\boldsymbol{v}}(h)$ and the corresponding estimated local Gaussian spectral densities $\widehat{f}_{\boldsymbol{v}}^{m}(\omega)$.

The computational cost for one single estimate of the local Gaussian correlation $\rho_{\boldsymbol{v}}(h)$, for a given lag $h$, a given bandwidth $\boldsymbol{v}$ and a given point $\boldsymbol{v}$, is usually not that high (depends on the sample size $n$). The computational cost does however quickly escalate when a huge combination of points $\boldsymbol{v}$, bandwidths $\boldsymbol{b}$ and large truncation level $m$ is used. It becomes even worse if it is of interest to produce pointwise confidence intervals, since it then will be necessary to have $R$ replicates of every configuration of these tuning parameters.

This implies that it for a practical investigation is natural to first do the computations on a single sample, a few bandwidths $\boldsymbol{b}$ and a wide range of points $\boldsymbol{v}$. The truncation level

$m$ could in this initial investigation probably be rather low, e.g. $m = 30$, since the key observation is that it is differences between the $m$-truncated ordinary and local Gaussian spectra that can reveal the presence of non-Gaussian dependency structures in the sample.

The next step of the investigation is the inspection of the heatmap- and distance-based plots of the estimates $\widehat{f}_{\boldsymbol{v}}^{m}(\omega)$, and from this it is then possible to figure out if there are some subset of the points $\boldsymbol{v}$ that it would be of particular interest to investigate further. If such points are identified, then it is possible to restrict another investigation to these points, and then perform e.g. $R = 100$ replicates in order to produce the pointwise confidence intervals.

This procedure was used in appendix G.4.3, where the aim of the investigation was to show that for a sufficiently large sample from the *local trigonometric* model used in section 3.3.2, it should be possible to detect the $C_1(t)$ component that only occurred with a probability of $p_1 = 0.05$. In this case a range of diagonal points $\boldsymbol{v}$ were selected from the lower tail, and one sample was used as the basis for the heatmap- and distance-based plots seen in fig. G.2. From this it was then easy to identify a suitable point $\boldsymbol{v}$ that could be used to create the plot in fig. G.3, where the pointwise confidence intervals also are present.

This kind of investigation is easy to reproduce for other samples, since the scripts in the `R`-package `localgaussSpec` can be modified in order to deal with similar investigations, cf. the discussion in appendix G for further details.

## E.2 Some comments regarding the bandwidth $b$

The bandwidth $\boldsymbol{b} = (.5, .5)$ used as default in section 3 of the main part was selected based on the fact that $b = .5$ is quite close to the value obtained when the formula $b \approx 1.75n^{-1/6}$ was given the value $n = 1974$ (the length of the `dmbp`-data). This formula, due to Håkon Otneim, is based on an empirical comparison with a cross-validation bandwidth algorithm used in Otneim and Tjøstheim [2017], and it has been applied here even though it originates from a bandwidth-selection algorithm aimed at computing density estimates based on the one-free-parameter local Gaussian approximation employed in that paper.

There does exist a leave-one-out cross-validation algorithm for the selection of the bandwidth to be used when estimating the local Gaussian correlation based on independent observations, see Berentsen and Tjøstheim [2014, Section 3.4] for details. However, the estimation of the local Gaussian spectral density $f_{\boldsymbol{v}}^{m}(\omega)$ requires the estimation of $m$ different local Gaussian autocorrelations $\rho_{\boldsymbol{v}}(h)$, and such cross-validation algorithms then becomes quite time consuming[13] — in particular if it in addition is necessary to use bootstrapping in order to obtain pointwise confidence intervals for the estimates. Moreover, it may be a

[13] Tests were performed to see if it might be possible to only use the bandwidth-algorithm for the case $h = 1$, and then let the higher lags inherit the estimated bandwidth — but it turned out that that

bit questionable to apply an algorithm developed for independent observations in a time series setting. In particular, the leave-one-out cross-validation has some flaws if the aim is model selection based upon dependent data, see Burman et al. [1994]; Racine [2000]; Shao [1993], where the concepts leave-$n_\nu$-out cross-validation, $h$-block cross validation, and $hv$-block cross-validation were introduced as better tools for the dependent case.

# Appendix F: Regarding sampling and resampling

This section will discuss sampling related issues, both with regard to the parametric and the nonparametric cases. Details related to the trivial case of sampling from parametric models are given in appendix F.1. Appendix F.2 discusses the approach based on parametric bootstrapping, which can be of interest in order to see if samples from a model fitted to a given data-set have the same dependency structure as the original data. This section includes a plot similar to one of the diagnostic plots used in Birr et al. [2019], in which points $\boldsymbol{v}$ both on and off the diagonal have been used in the investigation.

Nonparametric and model free bootstrap strategies are discussed in appendices F.3 and F.4, and it is there seen that a slightly adjusted version of the block bootstrap, cf. theorem F.4 on page 75, can be a useful resampling strategy for the estimators that are used to find the local Gaussian spectral densities.

A sensitivity analysis of the block length argument $L$ (used in the adjusted resampling algorithm) is given in appendix F.5, and a few additional comments related to problematic issues with the initial approach are given in appendix F.6.

## F.1 Simulations from a parametric model

Simulations are trivial for parametric time series models, since new independent samples (of the same length $n$) can be made directly from the model. The estimates of $f_{\boldsymbol{v}}^m(\omega)$ (for the specified values of $m$ and $\boldsymbol{b}$) are then computed for each of these samples, the mean of the resulting estimated spectra is used as the proxy for the true spectra, whereas pointwise confidence intervals are constructed directly from the collection of estimated spectra.

## F.2 Parametric bootstrap and local sanity-testing of models

A parametric bootstrap approach can be used to investigate models fitted to real data, and this is e.g. used in Birr et al. [2019]. The idea behind the parametric bootstrap is that a parametric model first is fitted to the original sample, and then that fitted model is used

assumption was not a viable one. In particular, the bandwidths estimated for the higher lags did not need to be close to the one estimated for the first lag.

when resampling — which implies that the second step in this procedure is identical to the one described in appendix F.1.

This approach can be used to perform a local sanity-test of the fitted model, since it becomes possible to identify points/frequencies with a clear mismatch between the local structures detected in the original sample and those seen in samples from the fitted model. The plot presented in fig. F.1, which is similar in structure to one of the plots in Birr et al. [2019], shows how such a comparison can be performed for the `dmbp`-data and the apARCH(2, 3)-model that was seen in figs. 9 and 11 in section 3.4.

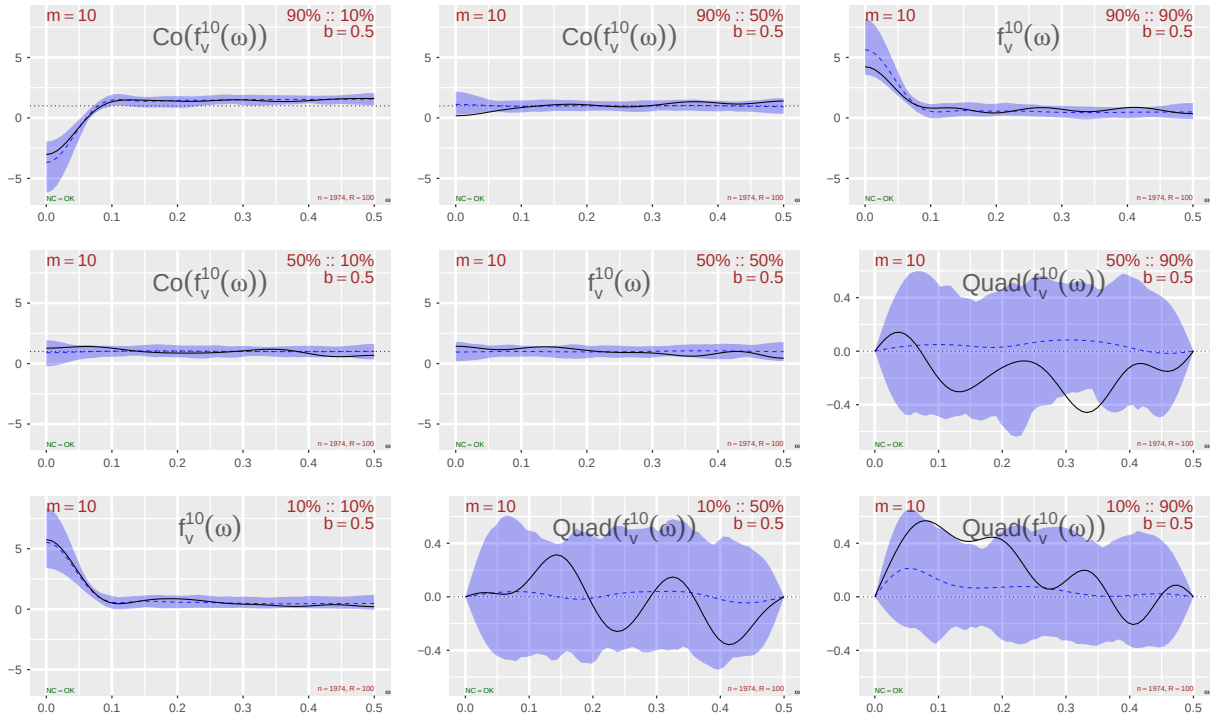


Figure F.1: The estimates of $f_{\boldsymbol{v}}^m(\omega)$ based on the `dmbp`-data (solid lines) have been superimposed on the corresponding estimates based on samples from the fitted apARCH(2, 3)-model. The off-diagonal points $\boldsymbol{v}$ give complex-valued $f_{\boldsymbol{v}}^m(\omega)$, see main text for explanation.

The key idea in fig. F.1 is that estimates of $f_{\boldsymbol{v}}^m(\omega)$ based on the original sample can be superimposed on the plots based on parametric bootstrapping from the fitted model, and this makes it easy to compare them.

Nine different points $\boldsymbol{v} = (v_1, v_2)$ are considered in fig. F.1, and these are based on the combinations that can be created when $v_1$ and $v_2$ varies over the 10%, 50% and 90% percentiles of the standard normal distribution. The corresponding plots are ordered in a grid in accordance with the position of these nine points in the plane, as can be seen by the information about $\boldsymbol{v}$ in the upper right corner of the respective plots.

The estimates of $f_{\boldsymbol{v}}^m(\omega)$ for the three diagonal points are real-valued, and this is thus in essence the same plots that was seen in fig. 11 — but the information about the global spectrum has been removed and the solid lines from fig. 9 have been added to the plots.

The estimates $f_{\boldsymbol{v}}^m(\omega)$ are complex-valued for the six off-diagonal points, and in this case the `R`-package `localgaussSpec` follows the convention used for the complex-valued cross-spectra, viz. $\mathrm{Co}(f_{\boldsymbol{v}}^m(\omega)) = \mathrm{Re}(f_{\boldsymbol{v}}^m(\omega))$ and $\mathrm{Quad}(f_{\boldsymbol{v}}^m(\omega)) = -\mathrm{Im}(f_{\boldsymbol{v}}^m(\omega))$.

The off-diagonal points are symmetric around the diagonal, i.e. both $\boldsymbol{v} = (v_1, v_2)$ and its diagonal reflection $\breve{\boldsymbol{v}} = (v_2, v_1)$ are present. It is the case that $f_{\boldsymbol{v}}(\omega) = \overline{f_{\breve{\boldsymbol{v}}}(\omega)}$, cf. lemma 2.3(b), so it is sufficient to plot $\mathrm{Co}(f_{\boldsymbol{v}}^m(\omega))$ on one side of the diagonal and $\mathrm{Quad}(f_{\boldsymbol{v}}^m(\omega))$ on the other side.

Finally, the same scale is used for all plots showing real values, whereas another scale is used for the plots related to the imaginary parts. This distinction is natural since the scale needed for the imaginary part can be much smaller, as can be seen in fig. F.1.

A comparison of the dashed and solid lines in fig. F.1 can now be used to see if there might be any faults with the apARCH(2, 3)-model that was fitted to the `dmbp`-data. The plots related to the real parts does not give any indications that something is off, with a possible minor exception near $\omega = 0$ for the point at the upper tail (as also observed in section 3.4.3 in the main part). The plots related to the imaginary parts might (when seen isolated) imply that the model did not catch all of the dependency structure — but it is here important to keep in mind that different scales are used for the two groups of plots, and as such it seems natural that a good match at the dominating scale might be accompanied with a more messy situation at the other scale.

It seems natural to conclude that the selected apARCH(2, 3)-model performs rather well, which is as expected since it was one of the better models from a testing procedure that tried out several thousand different variations of the GARCH-type models implemented in the `rugarch`-package.

A final comment to fig. F.1: Note the shape seen for the points $\boldsymbol{v} = (v_1, v_2)$ on the outer tails of the anti-diagonal, viz. when $v_1$ corresponds to the 10% percentile and $v_2$ to the 90% percentile (or vice versa). For these points, $\mathrm{Co}(f_{\boldsymbol{v}}^m(\omega))$ does have a deep trough near $\omega = 0$, which is rather natural since in a volatile situation it can be the case that a large decrease is followed by a somewhat larger increase (like a 'Sucker Rally' in the stock-market).

## F.3 Nonparametric bootstrapping techniques

This section will first explain why the block bootstrap could be a reasonable resampling technique for a statistic like the $m$-truncated estimates of the local Gaussian spectra. It will then be seen that after all there are some issues with the block bootstrap for the present case, and that motivates the quest for a slightly modified resampling strategy.

Technical details related to the bootstrap and block bootstrap are collected in appendix F.3.1, whereas appendix F.3.2 discuss some problems related to edge-effects between the blocks in the resampled time series. Appendix F.3.3 discuss one potential solution to the edge-effect issue, and explains why this approach was discarded for the investigation performed in the present paper. Appendix F.3.4 presents the ideas behind the block-of-blocks bootstrap (where edge-effects does not occur), and it explains why a direct application of that method might not be an optimal approach when the statistic of interest is computed by means of an algorithm that contains a kernel function.

**Justification for the block bootstrap:** First of all, recall from algorithm 2.5(c) (in the main part) that the $m$-truncated estimates $\widehat{f}_{\boldsymbol{v}}^{m}(\omega)$ of the local Gaussian spectral densities $f_{\boldsymbol{v}}(\omega)$, are constructed as follows:

$$\widehat{f}_{\boldsymbol{v}}^{m}(\omega) := 1 + \sum_{h=1}^{m} \lambda_{m}(h) \cdot \widehat{\rho}_{\breve{\boldsymbol{v}}}(h|\boldsymbol{b}_{h}) \cdot e^{+2\pi i\omega h} + \sum_{h=1}^{m} \lambda_{m}(h) \cdot \widehat{\rho}_{\boldsymbol{v}}(h|\boldsymbol{b}_{h}) \cdot e^{-2\pi i\omega h}, \tag{F.1}$$

where the point $\breve{\boldsymbol{v}} = (v_2, v_1)$ is the diagonal reflection of $\boldsymbol{v} = (v_1, v_2)$, and $\boldsymbol{b}_{h}$ is the bandwidth-vector used for the lag-$h$ pairs (the $\boldsymbol{b}_{h}$ will henceforward be dropped from the notation).

Note that the estimates $\widehat{\rho}_{\breve{\boldsymbol{v}}}(h)$ and $\widehat{\rho}_{\boldsymbol{v}}(h)$, for $h = 1, \ldots, m$, and also the $m$-truncated estimate $\widehat{f}_{\boldsymbol{v}}^{m}(\omega)$, all are estimated by a local likelihood approach — and the asymptotic properties of these estimates were developed in the present paper using the procedure from Klimko and Nelson [1978], cf. the discussion in appendix B.1.

A statistic obtained from the Klimko-Nelson procedure was explicitly mentioned by Künsch as an example for which the block bootstrap method would be applicable, cf. Künsch [1989, Example 2.4, p. 1219-20], and a resampling based on the block bootstrap was thus initially used for the construction of the pointwise confidence intervals for the `dmbp`-example seen in fig. 9.

Comments received during the review-process initiated an investigation of the following problem: Estimates based on the block bootstrap method can suffer from edge-effect noise when it is used on smaller sample sizes, cf. the discussion in appendix F.3.2. This motivated an investigation of possible replacements, that in the end lead to the slightly adjusted version of the block bootstrap given in appendix F.4, see theorem F.4 on page 75.

### F.3.1 The bootstrap and the block bootstrap

The bootstrap introduced in Efron [1979] use sampling with replacement from an i.i.d. sample $\{X_i\}_{i=1}^n$ to create a collection of $B$ bootstrapped samples $\{\{X_{i:b}^*\}_{i=1}^n\}_{b=1}^B$, and then a nonparametric estimator of the variance of a statistic $T_n := T(X_1, \dots, X_n)$ can be computed from the estimates in $\{T_{n:b}^*\}_{b=1}^B$, where $T_{n:b}^* := T(X_{1:b}^*, \dots, X_{n:b}^*)$. The block bootstrap introduced in Künsch [1989] enables a similar investigation to be performed when the statistic $T_n$ is computed on a set of observations $\{X_i\}_{i=1}^n$ from a stationary process, and in this case the resampled sets $\{X_{i:b}^*\}_{i=1}^n$ are created by the following procedure: (1) Create the set of $L$-sized blocks of consecutive observations from $\{X_i\}_{i=1}^n$, i.e. $\{\boldsymbol{Y}_i\}_{i=1}^{n-(L-1)}$, where $\boldsymbol{Y}_i = \big(X_i, \dots, X_{i+(L-1)}\big)$. (2) Sample with replacement $\lceil n/L \rceil$ of these blocks, to obtain a set $\{\boldsymbol{Y}_{i:b}^*\}_{i=1}^{\lceil n/L \rceil}$. (3) Concatenate the selected blocks to one block of size $\lceil n/L \rceil \cdot L$, and truncate it at length $n$ to obtain the desired resampled version $\{X_{i:b}^*\}_{i=1}^n$.

Künsch [1989] lists a wide range of different types of statistics that can be based on $\{X_{i:b}^*\}_{i=1}^n$, and it is for the purpose of the present paper of particular interest to note that statistics based on the Klimko-Nelson procedure is specifically mentioned as a case, which as mentioned above is the case for the estimators in this paper.

### F.3.2 Corrupt tuples and edge-effect noise for the block bootstrap

A problematic issue with the block bootstrap is that it will introduce a bit of *edge-effect noise* into the estimation procedure. For example, if a time series $\{Y_t\}_{t=1}^n$ of length $n$ is given, then an estimate of $\rho_{\boldsymbol{v}}(h)$ will be based on the bivariate set $\mathcal{Y}_h := \big\{\big(Y_{t+h}, Y_t\big)\big\}_{t=1}^{n-h}$ of size $n-h$. When the block bootstrap is used with some block length $L$, then there will be a resampled sequence $\{Y_t^*\}_{t=1}^n$ and the idea is that an estimate of $\rho_{\boldsymbol{v}}(h)$ now should be computed based on the bivariate set $\mathcal{Y}_{h:L}^* := \big\{\big(Y_{t+h}^*, Y_t^*\big)\big\}_{t=1}^{n-h}$.

However, the set $\mathcal{Y}_{h:L}^*$ will contain *corrupt tuples* that do not exist in $\mathcal{Y}_h$, i.e. the first and second component of $\big(Y_{t+h}^*, Y_t^*\big)$ can belong to different blocks, and this will add a bit of *edge-effect noise* into the estimation process. The edge-effect noise is negligible in the asymptotic situation (very large sample sizes $n$ and large block lengths $L$), but it can make an impact when smaller samples are investigated.

For the present paper, it is of particular interest to consider the amount of corrupt tuples that occur when the block bootstrap is used on the `dmbp`-data ($n = 1974$ unique observations, i.e. no ties). The plots in fig. 9 used the truncation level $m = 10$ for $f_{\boldsymbol{v}}^m(\omega)$, and it is thus natural to focus on the estimation of $\rho_{\boldsymbol{v}}(h)$ for $h = 1, \dots, 10$.

It is easy to see that the expected number $\mathcal{E}^*_{h:L}$ of corrupt tuples in $\mathcal{Y}^*_{h:L}$ to a close approximation[14] will be a simple formula of the number of blocks $q := \lceil n/L \rceil$ and the length $r := n - (q-1) \cdot L$ of the last block, i.e.

$$\mathcal{E}^*_{h:L} \approx \begin{cases} h \cdot (q-1) & h \leq r \leq L \\ h \cdot (q-2) + r & 1 \leq r < h. \end{cases} \tag{F.2}$$

A total of $n-h$ tuples $\left(Y^*_{t+h}, Y^*_t\right)$ are included in $\mathcal{Y}^*_{h:L}$, and the expected fraction of corrupt tuples is thus given by $\mathcal{E}^*_{h:L}/(n-h)$. It is enlightening to compute the expected fractions of corrupt tuples for the `dmbp`-data for the two block lengths $L = 25$ and $L = 100$, and the results (given as percentages) are listed in table 2.

| $L \setminus h$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 25 | 4.0% | 7.9% | 11.9% | 15.8% | 19.8% | 23.8% | 27.8% | 31.7% | 35.7% | 39.7% |
| 100 | 1.0% | 1.9% | 2.9% | 3.9% | 4.8% | 5.8% | 6.8% | 7.7% | 8.7% | 9.7% |

Table 2: The expected fraction of corrupt tuples when $\rho_{\boldsymbol{v}}(h)$ are estimated from block bootstrap replicates of the `dmbp`-data ($n = 1974$), when $L \in \{25, 100\}$ and $h \in \{1, \ldots, 10\}$.

It is evident, based on table 2, that the expected fraction of corrupt tuples can become rather large when $\rho_{\boldsymbol{v}}(h)$ is estimated for high lags $h$. The problem for estimates of $f^m_{\boldsymbol{v}}(\omega)$ is slightly reduced since the estimates $\widehat{\rho}_{\boldsymbol{v}}(h)$ are weighted with the lag-window functions $\lambda_m(h)$ when $\widehat{f}^m_{\boldsymbol{v}}(\omega)$ is computed, which implies that the estimates $\widehat{\rho}_{\boldsymbol{v}}(h)$ suffering from the highest levels of edge-effect noise do not contribute that much to the final result.

Table 2 indicates that it could be of interest to find an adjusted resampling technique, preferably one that completely (or at least partially) removes the corrupt tuples from the estimation algorithm. Two different approaches that completely avoids the corrupt tuples are presented in appendices F.3.3 and F.3.4, but there are some issues with these two methods that make them less interesting to implement.

It is however possible to reduce the number of corrupt tuples by slightly tweaking the way the block bootstrap algorithm is used when applied to smaller sample sizes. The key idea is to move the primary focus to the indices of the original sample, and then apply a simple adjustment that selects the $h$-lag pairs in a manner that is more in line with the way these pairs would have been selected if the methods from appendices F.3.3 and F.3.4 had been used. The technical details are given in appendix F.4, see in particular theorem F.4.

[14] It is possible that two neighbouring blocks can join perfectly (no edge-effect noise), so the correct formula for the expected number of corrupt tuples is slightly less than the numbers given in eq. (F.2), but this level of precision is not needed for the present discussion.

The corrupt tuples do not disappear with the adjusted resampling strategy from theorem F.4, but the expected fraction of such tuples (for a given combination of sample size $n$, block length $L$ and lag $h$) is significantly lower than those seen in table 2. It can e.g. be seen from table 3 (page 80) that for the $h = 10$ case it will be a reduction from 39.7% to 0.11% when $L = 25$, and a reduction from 9.67% to 0.028% when $L = 100$.

### F.3.3 A 'natural' solution to the edge-effect issue?

Obviously, if the aim of the investigation is restricted to $\widehat{\rho}_{\boldsymbol{v}}(h)$ for a single value of $h$, then it is trivial to completely avoid the problem of corrupt tuples in $\mathcal{Y}^*_{h:L}$. The solution in that case would simply be to realise $\mathcal{Y}_h$ as a sample from a bivariate time series, and then apply the block bootstrap method on $\mathcal{Y}_h$ instead of the original sample. The situation becomes a bit more complicated when it is necessary to estimate $\widehat{\rho}_{\boldsymbol{v}}(h)$, for $h = 1, \ldots, m$, since an approach where each of these estimates are computed from its own $\mathcal{Y}_h$ might fail to capture some of the temporal dependency structure from the original sample $\{Y_i\}_{i=1}^n$.

The temporal dependency structure between $\widehat{\rho}_{\boldsymbol{v}}(h)$ will be taken care of if the estimation of $\{\rho_{\boldsymbol{v}}(h)\}_{h=1}^m$ is based on (the relevant parts of) the $(m+1)$-tuples in the derived time series $\mathcal{Y}_{\bar{m}} = \left\{\left(Y_{i+m}, \ldots, Y_{i+1}, Y_i\right)\right\}_{i=1}^{n-m}$, but this approach is slightly wasteful since the estimation of $\rho_{\boldsymbol{v}}(h)$ for an $h < m$ in this case discards the last $m - h$ observations that would have been used if the estimate had been based on $\mathcal{Y}_h$ instead. The effect of this wastefulness will of course not be severe when a large sample is investigated, but it is present.

Moreover, this approach implies that the estimates of $\rho_{\boldsymbol{v}}(h)$, for $1 \leq h \leq m$, will depend on the selected value $m$. For a strict regime of reproducibility, like the one implemented in the `R`-package `localgaussSpec` , this implies that everything must be recomputed if the initial truncation level $m_1$ is changed to $m_2$. The computational cost related to the estimate of $\rho_{\boldsymbol{v}}(h)$ (for a fixed point $\boldsymbol{v}$ and a fixed bandwidth $\boldsymbol{b}$) is usually not that high, but a local Gaussian investigation will typically involve a wide range of lags $h$, many points $\boldsymbol{v}$, different values of the bandwidth $\boldsymbol{b}$, and a huge number of replicates. This implies that the number of cases to recompute might increase to the tens of thousands, which makes the 'resampling from $\mathcal{Y}_{\bar{m}}$ seen as an $(m+1)$-variate time series' approach far from desirable to implement.

The new estimation algorithm introduced in appendix F.4 are inspired by the resampling from tuples outlined above, and for the cost of a tiny percentage of edge-effects it will completely avoid the problematic issues mentioned. In particular: The estimation of the local Gaussian autocorrelations $\rho_{\boldsymbol{v}}(h)$ will use all the available information in $\mathcal{Y}_h$, and the estimated values $\widehat{\rho}_{\boldsymbol{v}}(h)$ will be the same regardless of the value of the truncation level $m$.

**The role of the block length $L$ when resampling from $\mathcal{Y}_{\bar{m}}$:** The discussion in appendix F.4 will reveal that the block length $L$ plays a different role when the block bootstrap is used on the $(m+1)$-variate tuples in $\mathcal{Y}_{\bar{m}}$, since both $m$ and $L$ then contribute to the capturing of the desired dependency structure. This is different from the situation seen when the ordinary block bootstrap is used on $\{Y_t\}_{t=1}^n$, since then it only is the block length $L$ that decides to what extent the temporal dependency structure of the original sample is preserved in the resampled data $\{Y_t^*\}_{t=1}^n$. In particular, a too short block length will simply destroy all of the dependency structure that it is of interest to investigate.

The situation changes when the block bootstrap is used on $\mathcal{Y}_{\bar{m}}$ (regarded as an $(m+1)$-dimensional time series), since it for some estimators then might be the case that even a very short block length $L$ can give decent results (in particular for an estimator that focus solely on the content captured in the $(m+1)$-variate tuples). For example: If $L = 1$, then the block bootstrap used on $\mathcal{Y}_{\bar{m}}$ is equivalent to uniform sampling from the tuples in $\mathcal{Y}_{\bar{m}}$. For an estimator that does not care about the internal order of the resampled tuples, e.g. the local likelihood estimator used in this paper, it might then in fact be sufficient to use such a short block length.

The block length argument $L$ is for this particular situation reduced to a tuning parameter that governs the expected number of times the different tuples occur in the resampled version of $\mathcal{Y}_{\bar{m}}$. A higher value of the block length $L$ will slightly reduce the fraction of tuples sampled from the start and the end of $\mathcal{Y}_{\bar{m}}$, whereas the majority of the tuples will have a tiny increase in the expected number of occurrences, cf. appendix F.4.3.

The reduction in the expected number of tuples sampled from the end of the time series can be of interest for the adjusted resampling strategy given in appendix F.4, since it will induce a corresponding reduction in the expected number of corrupt tuples, which is desirable since it removes some of the expected edge-effect noise from the estimation. See the discussion in appendix F.4 for further details.

### F.3.4 The block-of-blocks bootstrap

Another tuple-based bootstrapping approach that should be mentioned is the block-of-blocks bootstrap introduced in Politis and Romano [1992]. This method completely avoids the edge-effect issue that was mentioned for the block bootstrap, which makes it an interesting alternative to consider.

The key idea in the block-of-blocks bootstrap is that two levels of blocks are created, and resampling is made from the second level. The first level of blocks are created as follows: For a strictly stationary and weakly dependent $d$-variate time series $\{\boldsymbol{X}_i\}_{i=1}^n$, let $B_{i,m,L} := \big(\boldsymbol{X}_{(i-1)L+1}, \ldots, \boldsymbol{X}_{(i-1)L+m}\big)$. The block $B_{i,m,L}$ contains $m$ consecutive observations, and it can

be considered the result of a 'window' of width $m$ that is 'moving' at lags $L$ at a time. There are $Q = \lceil (n-m)/L \rceil$ of these blocks, and for each block a statistic $T_{i,m,L}$ is defined by a function $\phi_m : \mathbb{R}^{dm} \to \mathbb{R}$, i.e. $T_{i,m,L} := \phi_m\big(B_{i,m,L}\big)$. Note that the set $\big\{T_{i,m,L}\big\}_{i=1}^{Q}$ actually is a sample from a strictly stationary univariate time series (derived from the original time series through $\phi_m$), and note that the mean of $\big\{T_{i,m,L}\big\}_{i=1}^{Q}$, i.e. $\overline{T}_n := \frac{1}{Q}\sum_{i=1}^{Q} T_{i,m,L}$, gives an estimate of the true value of the statistic given by the aforementioned function $\phi_m$. It is thus of interest to do a block bootstrap on the sample $\big\{T_{i,m,L}\big\}_{i=1}^{Q}$ in order to investigate the properties of the estimator $\overline{T}_n$ — and this motivates the creation of the second level of blocks $\mathcal{B}_j$, which are created from $\big\{T_{i,m,L}\big\}_{i=1}^{Q}$ by means of a 'window' of width $L$ that is 'moving' at lags $h$ at a time: $\mathcal{B}_j := \big(T_{(j-1)h+1,m,L}, \ldots, T_{(j-1)h+L,m,L}\big)$ is constructed by taking $L$ consecutive observations from $\big\{T_{i,m,L}\big\}_{i=1}^{Q}$, and there are $q = \lceil (Q-L)/h \rceil$ of these blocks. Politis and Romano [1992, p. 1993] explain how sampling with replacement ($k$ times), followed by a concatenation, can be used to construct resampled sets $T_1^*, \ldots, T_{kL}^*$, and they give the required theoretical results that connects the mean $\overline{T}^*$ of this sample with the mean $\overline{T}_n$ — which thus gives the algorithm for the block-of-blocks bootstrapping.

The block-of-blocks bootstrap completely avoids the edge-effect problem that occurs when the block bootstrap is used, since the statistic of interest (given by the function $\phi_m$) are computed on the individual blocks $B_{i,m,L}$. This restriction to individual blocks can be an excellent idea for many statistics of interest, but it is a somewhat questionable approach for the estimates $\widehat{\rho}_{\boldsymbol{v}}(h)$ of the local Gaussian autocorrelations. The reason for this is that the bandwidth argument $\boldsymbol{b}$ in the kernel function $K_{\boldsymbol{b}}(\boldsymbol{w}-\boldsymbol{v})$ must be much larger if the estimation algorithm is to be used on only a subset of the observations — and the local structures of interest might then not be detected at all.

It would of course be of interest to implement the block-of-block bootstrap for the estimates of the local Gaussian spectra if very large samples are encountered, i.e. when the individual blocks contains several thousand consecutive observations — but for shorter samples (like the `dmbp`-example) it seems better to use something else.

## F.4 A slightly adjusted resampling algorithm

This section will present a minor adjustment of the ordinary block bootstrap. The adjusted approach will by construction return the same results as those obtained from the ordinary block bootstrap when the sample size $n$ and the block length $L$ are large. The situation is different for smaller sample sizes, since the adjusted approach then will remove the majority of the corrupt tuples that adds edge-effect noise into the estimation of the local Gaussian autocorrelations $\rho_{\boldsymbol{v}}(h)$.

Note that this adjusted resampling strategy is designed to take care of statistics that are constructed from pairs $\left(Y_{t+h}, Y_t\right)$, and it does this by mimicking key features of the optimal resampling strategy described in appendix F.3.3. In contradistinction to the adjusted block bootstrap, the ordinary block bootstrap is not restricted to statistics based on pairs $\left(Y_{t+h}, Y_t\right)$, nor is it specially designed for such a case.

The block length $L$ plays a different role when the resampling is done on $(m+1)$-tuples, and it can be considered as a tuning parameter that governs the expected number of times the different tuples will occur in the resampled set, cf. the discussion at the end of appendix F.3.3. The sensitivity analysis of the block length $L$ in appendix F.5 indicates that the selection of $L$ should not be a problematic issue when the samples are large enough.

### F.4.1 A toy example to illustrate the principle

It will be a bit easier to digest the definitions and the algorithm that are given later on in this section, if a simple toy-example is investigated first: Consider a situation with a time series having five unique observations $Y_1, Y_2, Y_3, Y_4, Y_5$ and assume that there is an interest for an estimate based on the four lag-1 tuples in $\mathcal{Y}_1 = \left\{\left(Y_{t+1}, Y_t\right)\right\}_{t=1}^4$. If a block bootstrap with block length $L = 2$ is used, the resampled time series might e.g. look like $Y_1^* = Y_4, Y_2^* = Y_5, Y_3^* = Y_3, Y_4^* = Y_4, Y_5^* = Y_2$, and the corresponding set of lag-1 tuples would be $\mathcal{Y}_{1:2}^* = \left\{\left(Y_{t+1}^*, Y_t^*\right)\right\}_{t=1}^4$. It is easy to see that $\mathcal{Y}_{1:2}^*$ in this case will contain the two corrupt tuples $(Y_3, Y_5)$ and $(Y_2, Y_4)$, i.e. tuples that are not present in $\mathcal{Y}_1$.

The key idea in the adjusted algorithm is to move the focus to the indices of the original sample, i.e. $1, 2, 3, 4, 5$, and then use the block bootstrap to sample from these. The resampled set of indices for the example above would be $4, 5, 3, 4, 2$, and from these it is possible to construct the *cyclically* $h = 1$ *shifted* set of indices $5, 1, 4, 5, 3$. The method is simply to add the lag $h = 1$ to all the resampled indices — and to start back on 1 if a value exceeds $n = 5$. The four desired lag-1 tuples $\mathcal{Y}_{1:2}^\sharp = \left\{\left(Y_{t+1}^\sharp, Y_t^\sharp\right)\right\}_{t=1}^4$ are now created by using the resampled set of indices in the $Y_t^\sharp$-component, whereas the cyclically $h = 1$ shifted indices are used for the $Y_{t+h}^\sharp$-component. This results in the following four tuples, $\mathcal{Y}_{1:2}^\sharp = \{(Y_5, Y_4), (Y_1, Y_5), (Y_4, Y_3), (Y_5, Y_4)\}$, and it is easy to see that the only corrupt tuple in $\mathcal{Y}_{1:2}^\sharp$ is $(Y_1, Y_5)$. Note: It could in principle now also be added a fifth tuple $(Y_3, Y_2)$ to $\mathcal{Y}_{1:2}^\sharp$, but that is not of interest since there are only four tuples in $\mathcal{Y}_1$.

The adjusted resampling algorithm is thus quite simple in structure, and it only needs to be formalised. This is taken care of in definitions F.1 to F.3 and theorem F.4.

It is easy to compute the expected number of corrupt tuples in $\mathcal{Y}_{h:L}^\sharp$ for a given combination of sample size $n$, lag $h$, and block length $L$, and this is done in lemma F.5 in

appendix F.4.4. It can from this easily be seen how much the edge-effect noise is reduced for estimates based on the `dmbp`-data, cf. table 3 on page 80.

### F.4.2 Three definitions and one algorithm

**Definition F.1.** *For $n$ and $i$ positive integers, and $h$ a non-negative integer, define the new index $\mathcal{M}(i,h;n)$ as follows:*

$$\mathcal{M}(i,h;n) \coloneqq 1 + [(i+h-1) \mod n] = (i+h) - n \cdot \left\lfloor \frac{i+h-1}{n} \right\rfloor \tag{F.3}$$

The result of $\mathcal{M}(i,h;n)$ will always be a number in the set $\{1,\ldots,n\}$, and $\mathcal{M}(i,0;n) = i$ when $i \leq n$.

**Definition F.2.** *For fixed positive integers $m$ and $n$, with $m < n$, and any starting index $i \in \{1,\ldots,n\}$, define the $(m+1)$-tuple $\mathfrak{M}(i;m,n)$ as follows:*

$$\mathfrak{M}(i;m,n) \coloneqq (\mathcal{M}(i,m,n),\ldots,\mathcal{M}(i,1,n),i) \tag{F.4}$$

The result of $\mathfrak{M}(i;m,n)$ will be referred to as an $(m+1)$-variate tuple of indices. It will have the desirable form $(i+m,\ldots,i+1,i)$ when $i \leq n-m$. The result will be *cyclically shifted* when $i \in \{n-m+1,\ldots,n\}$, i.e. the indices will in that case have the form $(\mathcal{M}(i,m,n),\ldots,1,n,\ldots,i)$. Note that it is trivial to tweak the definition of $\mathfrak{M}(i;m,n)$, if only a subset of the resulting indices is required. This is e.g. the case for the indices needed when estimating $\widehat{\rho}_v(h)$, where it only is the bivariate pairs $(\mathcal{M}(i,h,n),i)$ that it is of interest to consider.

**Definition F.3.** *For a sample $\{Y_i\}_{i=1}^n$ of length $n$, an integer $m < n$ and any starting index $i \in \{1,\ldots,n\}$, use the indices from $\mathfrak{M}(i;m,n)$ to define an $(m+1)$-variate tuple $\boldsymbol{Y}(i;m,n)$ as follows:*

$$\boldsymbol{Y}(i;m,n) \coloneqq \left(Y_{\mathcal{M}(i,m,n)},\ldots,Y_{\mathcal{M}(i,1,n)},Y_i\right) \tag{F.5}$$

*The resulting tuple will be referred to as 'desirable' when $i \leq n-m$, whereas it will be referred to as 'corrupt' when $i \in \{n-m+1,\ldots,n\}$.*

If a starting index $i$ is selected randomly from $\{1,\ldots,n\}$, then there is a probability of $p = \frac{n-m}{n}$ that the tuple $\boldsymbol{Y}(i;m,n)$ will be desirable, and a probability of $1-p = \frac{m}{n}$ that the tuple will be corrupt.

With these definitions, it is now time to present the adjusted resampling algorithm.

**Algorithm F.4** (Circular index-based block bootstrap for tuples)**.**
*Given a sample* $\{Y_i\}_{i=1}^n$ *of length* $n$ *from a strictly stationary time series, and a statistic* $T_n$ *that is given as a function* $\varphi_n$ *of the* $(m+1)$*-variate set* $\mathcal{Y}_{\bar{m}} \coloneqq \big\{\big(Y_{i+m},\ldots,Y_{i+1},Y_i\big)\big\}_{i=1}^{n-m}$*, i.e.* $T_n \coloneqq \varphi_n(\mathcal{Y}_{\bar{m}})$*. For a given block length* $L$*, let* $q$ *be the number* $\lceil n/L\rceil$*, and define a resampled set* $\mathcal{Y}^\sharp_{\bar{m}:L}$*, and* $T_n^\sharp \coloneqq \varphi_n(\mathcal{Y}^\sharp_{\bar{m}:L})$*, as follows:*

*(a) Sample with replacement* $q$ *numbers* $n_1,\ldots,n_q$ *from the index set* $\{1,\ldots,n-(L-1)\}$*.*

*(b) For* $j\in\{1,\ldots,q\}$*, let* $\mathcal{I}^\sharp_{j:L}$ *be the* $L$*-sized tuple* $\big(n_j, n_j+1,\ldots,n_j+L-1\big)$*.*

*(c) Let* $\mathcal{I}_n^\sharp = (i_1^\sharp,\ldots,i_n^\sharp)$ *be the* $n$*-sized tuple that occurs when the* $q$ *tuples* $\mathcal{I}^\sharp_{1:L},\ldots,\mathcal{I}^\sharp_{q:L}$ *first are concatenated into one tuple, and then truncated at length* $n$*.*

*(d) Use the first* $n-m$ *indices from* $\mathcal{I}_n^\sharp$ *as starting indices, and let* $\mathcal{Y}^\sharp_{\bar{m}:L}$ *be given by*

$$\mathcal{Y}^\sharp_{\bar{m}:L} \coloneqq \big\{\boldsymbol{Y}(i_j^\sharp;m,n)\big\}_{j=1}^{n-m}. \tag{F.6}$$

*(e) Use the function* $\varphi_n$ *to define the estimate* $T_n^\sharp$*, i.e.* $T_n^\sharp \coloneqq \varphi_n(\mathcal{Y}^\sharp_{\bar{m}:L})$*.*

The index set $\mathcal{I}_n^\sharp$ from theorem F.4(c) is the same set of indices that would occur if the block bootstrap was used to obtain a resampled version $\{Y_i^*\}_{i=1}^n$ of the original sample $\{Y_i\}_{i=1}^n$. This implies (assuming reasonable values for $L$ and $m$) that the majority of the tuples in $\mathcal{Y}^\sharp_{\bar{m}:L}$ also will be present in $\mathcal{Y}^*_{\bar{m}:L} \coloneqq \big\{\big(Y^*_{i+m},\ldots,Y^*_{i+1},Y^*_i\big)\big\}_{i=1}^{n-m}$, where the latter is the one that would have been used to get an estimate $T_n^* \coloneqq \varphi_n(\mathcal{Y}^*_{\bar{m}:L})$ if the ordinary block bootstrap was used.

All the desirable tuples in $\mathcal{Y}^*_{\bar{m}:L}$ will also be contained in $\mathcal{Y}^\sharp_{\bar{m}:L}$, and it is easy to see, cf. similar discussion in appendix F.3.2, that the number of desirable $(m+1)$-variate tuples in $\mathcal{Y}^*_{\bar{m}:L}$ at least must be $(n-m\cdot\lceil n/L\rceil)/(n-m)$. This fraction converges towards 1, given reasonable assumptions with regard to how fast $L\to\infty$ and $m\to\infty$ when $n\to\infty$, which thus implies that the content of $\mathcal{Y}^\sharp_{\bar{m}:L}$ and $\mathcal{Y}^*_{\bar{m}:L}$ in essence coincide when $n\to\infty$ — and it is thus natural to anticipate that the asymptotic behaviour of the estimates $T_n^\sharp$ and $T_n^*$ should be quite similar.

As previously mentioned, the block bootstrap is viable for a statistic based on the Klimko-Nelson procedure, cf. Künsch [1989, Example 2.4, p. 1219-20], and it is thus in particular applicable when estimating the local Gaussian autocorrelations $\rho_{\boldsymbol{v}}(h)$ and the $m$-truncated local Gaussian spectra $f^m_{\boldsymbol{v}}(\omega)$. The previously mentioned overlap between $\mathcal{Y}^\sharp_{\bar{m}:L}$ and $\mathcal{Y}^*_{\bar{m}:L}$ indicates that the *circular index-based block bootstrap for tuples* from theorem F.4 also should be a viable alternative for the statistics of interest for the present paper.

### F.4.3 The block length $L$ and the expected content of $\mathcal{Y}^\sharp_{m:L}$

The purpose of the adjusted resampling strategy is to provide the required data $\mathcal{Y}^\sharp_{m:L}$, that can replace the $(m+1)$-variate tuples in $\mathcal{Y}_m$ when the pointwise confidence intervals are to

be estimated for the original estimate of $f_{\boldsymbol{v}}^{m}(\omega)$. A sensitivity analysis of the block length $L$ is included in appendix F.5, and it is thus of interest to add some comments about the effect the block length $L$ has on the expected content of $\mathcal{Y}_{m:L}^{\sharp}$.

It is with regard to this discussion of interest to point out that the temporal connection *between* the $(m+1)$-variate tuples in $\mathcal{Y}_{m}$ and $\mathcal{Y}_{m:L}^{\sharp}$ does not affect the resulting estimates of $f_{\boldsymbol{v}}^{m}(\omega)$. The reason for this is that the algorithm that estimates the local Gaussian autocorrelations $\rho_{\boldsymbol{v}}(h)$ only cares about the points in the plane that are defined by the bivariate lag-$h$ tuples, that again are derived from these $(m+1)$-tuples. To clarify: The temporal aspect is pivotal with regard to the construction of the $(m+1)$-variate tuples in $\mathcal{Y}_{m:L}^{\sharp}$, but the order does not matter anymore when these tuples first have been constructed.

The main detail of interest is thus to figure out the expected number of times the different tuples will occur in $\mathcal{Y}_{m:L}^{\sharp}$.

The first detail to note is that the content of $\mathcal{Y}_{m}$ and $\mathcal{Y}_{m:L}^{\sharp}$ correspond to *starting indices* given by $(n-m)$-tuples from the index-set $\{1,\ldots,n\}$. For $\mathcal{Y}_{m}$ it is simply the tuple $(1,\ldots,n-m)$, whereas it for $\mathcal{Y}_{m:L}^{\sharp}$ is the $n-m$ first indices from the tuple $\mathcal{I}_{n}^{\sharp}$ that was introduced in theorem F.4(c).

A brief inspection of items (a) and (b) of theorem F.4 reveals that $\mathcal{I}_{n}^{\sharp}$ is built from $q=\lceil n/L\rceil$ tuples $\mathcal{I}_{j:L}^{\sharp}=\big(n_{j},n_{j}+1,\ldots,n_{j}+L-1\big)$, where the index $n_{j}$ has been sampled uniformly from the index-set $\{1,\ldots,n-(L-1)\}$. The length of the $q-1$ first of these tuples are $L$, whereas the last tuple might be shorter since it has to be truncated to the length $r=n-(q-1)\cdot L$ in order for $\mathcal{I}_{n}^{\sharp}$ to have the length $n$.

The expected content of $\mathcal{Y}_{m:L}^{\sharp}$ is thus related to the expected number of times different starting indices $k$ will occur in $\mathcal{I}_{n}^{\sharp}$, which again is related to the probability that the building blocks $\mathcal{I}_{j:L}^{\sharp}$ contains $k$. The situation for the $q-1$ first of these building blocks is the simplest. The basic observation is that the event '$\mathcal{I}_{j:L}^{\sharp}$ contains $k$' is equivalent to '$n_{j}\leq k\leq n_{j}+L-1$', which can be rewritten as '$k-(L-1)\leq n_{j}\leq k$'. The number $n_{j}$ must lie in the index set $\{1,\ldots,n-(L-1)\}$, so this latter event is equivalent to '$1\vee(k-(L-1))\leq n_{j}\leq k\wedge(n-(L-1))$'. This implies that the probability that the $L$-length tuple $\mathcal{I}_{j:L}^{\sharp}$ contains $k$ can be written out as

$$\mathrm{P}\big(\mathcal{I}_{j:L}^{\sharp}\text{ contains }k,\text{ for }j=1,\ldots,q-1\big)=\begin{cases}\frac{k}{n-(L-1)} & 1\leq k<L\\ \frac{L}{n-(L-1)} & L\leq k<n-L\\ \frac{n-(k-1)}{n-(L-1)} & n-L\leq k\leq n.\end{cases}\tag{F.7}$$

The argument for the last block is similar, but the truncation to length $r$ implies that it can not contain any indices above the value $n-(L-r)$.

$$\mathrm{P}\big(\mathcal{I}^\sharp_{q:L} \text{ contains } k\big) = \begin{cases} \frac{k}{n-(L-1)} & 1 \leq k < r \\ \frac{r}{n-(L-1)} & r \leq k < n-L \\ \frac{n-(L-r)-(k-1)}{n-(L-1)} & n-L \leq k \leq n-(L-r) \\ 0 & n-(L-r) < k \leq n. \end{cases} \tag{F.8}$$

The expected number of occurrences of an index $k$ in the index set $\mathcal{I}^\sharp_n$ can be found by simply summing the expected number of occurrences in the $q$ building blocks $\mathcal{I}^\sharp_{j:L}$, and this is easy to find from eqs. (F.7) and (F.8). For the purpose of the sensitivity analysis in appendix F.5, it is sufficient to observe that the expected number of occurrences of an index $k$ that lies in the set $\{L,\ldots,n-L\}$ is given by $(q-1)\cdot\frac{L}{n-(L-1)}+1\cdot\frac{r}{n-(L-1)}$, and it follows from $r=n-(q-1)\cdot L$ that this is the number $\frac{n}{n-(L-1)}$.

This shows how the block length $L$ affects the expected number of times different indices $k$ occurs in $\mathcal{I}^\sharp_n$, which as mentioned above reveals the expected number of times the corresponding $(m+1)$ tuple will occur in $\mathcal{Y}^\sharp_{m:L}$. It is clear from the fraction $\frac{n}{n-(L-1)}$ that it for a large enough $n$ will be a rather negligible effect on these expectations when $L$ is modified from e.g. 10 to 69 (which is the case in appendix F.5).

### F.4.4 Edge-effect noise for the adjusted resampling algorithm

This section will investigate the edge-effect noise that occurs when the adjusted resampling algorithm is applied, and this will in appendix F.4.5 be used to check that the fraction of corrupt tuples becomes minuscule when this algorithm is used on the `dmbp`-data ($n=1974$ unique observations, i.e. no ties).

**Lemma F.5.** *Given a sample $\{Y_i\}_{i=1}^n$ from a continuous-valued time series, and the corresponding derived time series of $(m+1)$-tuples $\mathcal{Y}_{\overline{m}} \coloneqq \big\{\big(Y_{i+m},\ldots,Y_{i+1},Y_i\big)\big\}_{i=1}^{n-m}$. For a given block length $L>m$, let $q\coloneqq\lceil n/L\rceil$ be the number of blocks used in the construction of the resampled version $\mathcal{Y}^\sharp_{\overline{m}:L}$ (introduced in theorem F.4), and let $r\coloneqq n-L\cdot(q-1)$ be the length of the last block. Let $\mathcal{E}^\sharp_{\overline{m}:L}$ denote the expected number of corrupt tuples in $\mathcal{Y}^\sharp_{\overline{m}:L}$, i.e. tuples not occurring in $\mathcal{Y}_{\overline{m}}$. The number $\mathcal{E}^\sharp_{\overline{m}:L}$ is then given by the following formula:*

$$\mathcal{E}^\sharp_{\overline{m}:L} = \begin{cases} \frac{1}{2}(q-1)\frac{m(m+1)}{n-(L-1)} & m \leq r \leq L \\ \frac{1}{2}(q-2)\frac{m(m+1)}{n-(L-1)} + \frac{1}{2}\frac{r(r+1)}{n-(L-1)} & 1 \leq r < m \end{cases} \tag{F.9}$$

*Proof.* The continuity-requirement implies that there are no ties (as is the case for the `dmbp`-data). Further, there is no need to adjust the result for the possibility that a corrupt index-set (of length $m+1$) can concatenate observations from the two ends of $\{Y_i\}_{i=1}^n$ into a sequence that already exists as a sub-sequence of $\{Y_i\}_{i=1}^n$. To clarify: This requirement ensures e.g. that no proper tuple $\big(Y_{i+1}, Y_i\big)$ can be equal to $(Y_1, Y_n)$, so the formulas in eq. (F.9) are thus exact and not only approximate.

The blocks used in the construction of $\mathcal{Y}^\sharp_{\bar{m}:L}$ are uniquely identified by the starting indices given in $\mathcal{I}^\sharp_{j:L} = \big(n_j, n_j+1, \ldots, n_j+(L-1)\big)$, where the initial numbers $n_1, \ldots, n_q$ are sampled uniformly from $\{1, \ldots, n-(L-1)\}$. This implies that $\mathcal{E}^\sharp_{\bar{m}:L}$ can be expressed as the sum of the expected number of corrupt tuples in the individual blocks.

It was mentioned in definition F.3 that zero corrupt tuples would occur for a starting index in $\{n-(m-1), \ldots, n\}$, and it follows from this that a block will contain 0 corrupt tuples when $n_j \leq n-m-(L-1)$. This implies that the probability for 0 corrupt tuples in a block is given by $\frac{n-m-(L-1)}{n-(L-1)}$. It is easy to check that a starting tuple given by $n_j = n-(m-k)-(L-1)$ for some $k \in \{1, \ldots, m\}$ must correspond to a block that contains $k$ corrupt tuples, and each of these outcomes have the same probability $\frac{1}{n-(L-1)}$. It follows from this that the expected number of corrupt tuples in a block is given by $\frac{1}{2}\frac{m(m+1)}{n-(L-1)}$.

The expected number of corrupt tuples for the individual blocks can now be used to compute $\mathcal{E}^\sharp_{\bar{m}:L}$, i.e. the expected number of corrupt tuples in $\mathcal{Y}^\sharp_{\bar{m}:L}$. Note that only the first $n-m$ indices from $\mathcal{I}^\sharp_n$ are used in the computation of $\mathcal{Y}^\sharp_{\bar{m}:L}$, cf. theorem F.4(d), and that implies that any potential corrupt tuples from the last block will be discarded due to this truncation. The length $r$ of the last block will thus influence whether or not some potential corrupt tuples from the second to last block also might be removed in this truncation, and the formula for the expected number of corrupt tuples in $\mathcal{Y}^\sharp_{\bar{m}:L}$ must thus take the value of $r$ into account. By construction, $r$ will be a number in the set $\{1, \ldots, L\}$.

The case where $r \geq m$ is the simplest, since the truncation to length $n-m$ in this case does not affect the second to last block. The expected number of corrupt tuples in $\mathcal{Y}^\sharp_{\bar{m}:L}$ is thus simply the sum of the expected number of corrupt tuples from the $q-1$ first blocks, which gives the result $\mathcal{E}^\sharp_{\bar{m}:L} = \frac{1}{2}(q-1)\frac{m(m+1)}{n-(L-1)}$ when $r \geq m$.

The situation for the case $r < m$ is slightly more complicated. The effect of truncation to length $n-m$ will in this case completely eliminate the last block of $\mathcal{I}^\sharp_n$, and the second to last block will have its last $(m-r)$ indices removed. This implies that the highest possible number of corrupt tuples from block number $q-1$ is reduced from $m$ to $r$, which implies that the expected number of corrupt tuples from this block becomes $\frac{1}{2}\frac{r(r+1)}{n-(L-1)}$. The stated

result follows when this expected number is added together with the contribution from the $q-2$ first blocks, i.e. $\mathcal{E}^{\sharp}_{\overline{m}:L} = \frac{1}{2}(q-2)\frac{m(m+1)}{n-(L-1)} + \frac{1}{2}\frac{r(r+1)}{n-(L-1)}$ when $r < m$. □

The result in lemma F.5 is stated for $m+1$ tuples of the form $\left(Y_{i+m}, \ldots, Y_{i+1}, Y_i\right)$, but it is easy to see that the expected number of corrupt tuples remains the same if it is restated for bivariate lag-$m$ tuples $\left(Y_{i+m}, Y_i\right)$. This implies that the formula in eq. (F.9) can be used for the `dmbp`-data investigation given in the next section.

The continuity requirement in lemma F.5 was included in order to avoid additional technicalities in the proof, but the resulting expression for the expected number of corrupt tuples would for most cases remain the same even if some observations were repeated.

A minor warning should be added with regard to the corrupt tuples that actually do occur when the resampling strategy from theorem F.4 is used: The way the tuples $\left(Y^{\sharp}_{t+h}, Y^{\sharp}_{t}\right)$ is constructed implies that the corrupt tuples always will occur at the exact same positions. For example, the lag-1 corrupt tuple will always be the tuple $(Y_1, Y_n)$, the lag-2 corrupt tuples will always be from the set $\left\{\left(Y_1, Y_{n-1}\right), (Y_2, Y_n)\right\}$, and so on. In a worst case scenario, some of these tuples might be very close to the point $\boldsymbol{v}$ for which the local Gaussian autocorrelation $\rho_{\boldsymbol{v}}(h)$ is to be computed (this can easily be checked by plotting the relevant tuples). Given the low expected fraction of corrupt tuples, cf. table 3 in the next section, it seems likely that this effect should not turn out to be a too big problem.

### F.4.5 The `dmbp`-data and corrupt tuples for the adjusted resampling algorithm

It was seen in appendix F.3.2 that the ordinary block bootstrap could produce a high fraction of corrupt tuples when it was used on smaller samples. The `dmbp`-data ($n = 1974$) was used as a test case, and table 2 on page 69 listed the approximate fractions of corrupt tuples that was expected to occur in $\mathcal{Y}^{*}_{h:L} = \left\{\left(Y^{*}_{t+h}, Y^{*}_{t}\right)\right\}_{t=1}^{n-h}$ when $h \in \{1, \ldots, 10\}$ and $L = \{25, 100\}$. It is now of interest to create a similar table for the *circular index-based block bootstrap for tuples* from theorem F.4, in order to see to what extent the adjusted resampling strategy manages to reduce the expected number of corrupt tuples $\mathcal{Y}^{\sharp}_{h:L} = \left\{\left(Y^{\sharp}_{t+h}, Y^{\sharp}_{t}\right)\right\}_{t=1}^{n-h}$.

The counting formula from eq. (F.9) can, as mentioned after the proof of lemma F.5, be used for the present case of interest too. The length of the last blocks will for the two cases $L = 25$ and $L = 100$ respectively be 24 and 74, and this implies (since both of them are larger than $h = 10$), that it is the version $\frac{1}{2}(q-1)\frac{h(h+1)}{n-(L-1)}$ that should be used to find the expected number of corrupt tuples in $\mathcal{Y}^{\sharp}_{h:L}$ when $h \in \{1, \ldots, 10\}$. The data in table 2 was given as fractions of the total number of tuples $n-h$, and table 3 has thus used the same adjustment.

A comparison with table 2, with focus on the entries in the $h = 10$ column, shows that the numbers have been reduced from 39.7% to 0.112% when $L = 25$, and it has been a

| $L \setminus h$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 25 | 0.002% | 0.006% | 0.012% | 0.020% | 0.030% | 0.043% | 0.057% | 0.073% | 0.092% | 0.112% |
| 100 | 0.001% | 0.002% | 0.003% | 0.005% | 0.008% | 0.011% | 0.014% | 0.019% | 0.023% | 0.028% |

Table 3: The expected amount of corrupt tuples when $\rho_v(h)$ are estimated for the `dmbp`-data by the *circular index-based block bootstrap for tuples*, cf. theorem F.4.

reduction from 9.7% to 0.028% when $L = 100$. This implies that the edge-effect noise for the adjusted resampling strategy can be considered rather negligible, and it could also be the case that estimates based on $\mathcal{Y}^{\sharp}_{h:L}$ might be less sensitive to changes in the block length $L$, cf. the sensitivity analysis in appendix F.5.

The relation between the entries in tables 2 and 3 can be found by comparing the counting formulas for the expected number of corrupt tuples in $\mathcal{Y}^{*}_{h:L}$ and $\mathcal{Y}^{\sharp}_{h:L}$, cf. respectively eqs. (F.2) and (F.9), and this results in[15]

$$\frac{\mathcal{E}^{\sharp}_{h:L}}{\mathcal{E}^{*}_{h:L}} \approx \begin{cases} \frac{1}{2}\frac{h+1}{n-(L-1)} & h \leq r \leq L \\ \frac{1}{2}\frac{(q-2)\cdot h\cdot(h+1)+r\cdot(r+1)}{(n-(L-1))\cdot(h\cdot(q-2)+r)} & 1 \leq r < h. \end{cases} \tag{F.10}$$

It follows from eq. (F.10) that the $L = 25$ and $h = 10$ entry in table 3 is 0.282% of the corresponding value in table 2 — and it can similarly be seen that the same relation for the entry $L = 100$ and $h = 10$ is given by 0.293%.

## F.5 Sensitivity analysis: The block length $L$

The block length sensitivity for the adjusted resampling strategy from theorem F.4 will now be investigated — and the computations will, as for the other tuning parameters investigated in appendix D, be based on the `dmbp`-data.

The tool for this investigation will be the distance function $D$ that was defined in appendix D.1, i.e. the distance function inherited from the complex Hilbert space of Fourier series on the interval $\left[-\frac{1}{2}, \frac{1}{2}\right]$. This distance function does not reveal anything about the frequency-component of the cases under investigation, so it is also necessary to include a plot that focus on that aspect for a few of the block lengths $L$.

The block length $L$ takes integer values, and one possible way to gain some insight into the sensitivity of this argument is to use a sequence of box-plots to show the status for different values of $L$. This approach has been used in fig. F.2, where the panel at the top contains the results as $L$ increases in steps of 1 from $L = 10$ to $L = 69$.

[15] The result is only approximate since eq. (F.2) did not adjust for the possibility that neighbouring blocks in some rare cases could join perfectly (no edge-effect noise), but the actual fraction should be very close to the one given by the expressions in eq. (F.10).

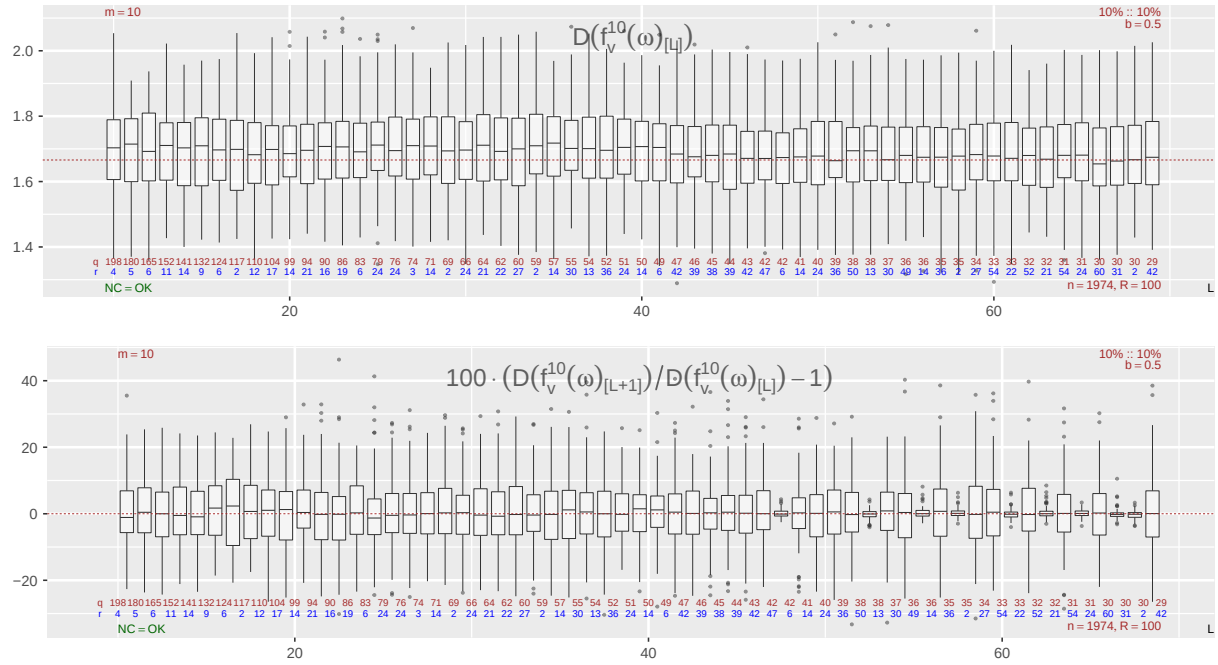


Figure F.2: Distance based box-plots for the investigation of the sensitivity of the block length $L$ for the adjusted resampling strategy from theorem F.4. The numbers in the two bottom rows show $q = \lceil n/L \rceil$ and $r = n - (q-1) \cdot L$, i.e. the number of blocks and the length of the last block.

**The panel at the top of fig. F.2:** A box-plot for the $D\big(f_v^{10}(\omega)_{[L]}\big)$-values (based on $R = 100$ replicates) is given for each block length $L$. A horizontal red dashed line has been added that shows the $D(f_v^{10}(\omega))$-value for the original sample. It can be seen that the medians of the box-plots tend to be slightly larger than the horizontal line that corresponds to the value based on the original sample, they seem to approach the line as $L$ increases, but these medians are based on $R = 100$ replicates — and another realisation might thus look slightly different. It does not seem to be any pattern here with regard to how these box-plots changes when $L$ increases.

**The panel at the bottom of fig. F.2:** These box-plots shows the percent-wise changes in the distances when the block length goes from $L$ to $L + 1$, and everything else

is kept identical, i.e. $100 \cdot \big(D\big(f_v^{10}(\omega)_{[L+1]}\big)/D\big(f_v^{10}(\omega)_{[L]}\big) - 1\big)$. This is possible to do since the reproducibility setup enables a tracking for each individual realisation.

A horizontal red dashed line has been added at 0, and it is clear that the median-part of these box-plots are quite close to this horizontal line. It can also be observed that some of these box-plots are more compact than the other ones, and a simple investigation of the numbers given at the bottom of the plots reveals that this phenomenon occurs when an increase from $L$ to $L+1$ does not reduce the number of blocks that are needed, i.e. they occur when $\lceil n/L \rceil = \lceil n/(L+1) \rceil$.

For the individual bootstrapped time series, this indicates that the changes are minimal when the number of blocks remains the same — whereas the changes are much larger when the increase of $L$ triggers a reduction in the number of blocks. However, as is evident from an inspection of the panel at the top of fig. F.2, this effect is only on the level of the individual replicates, and it is averaged away when a collection of replicates is considered.

Note that the effect noticed in the bottom panel of fig. F.2 also is present for the global spectral densities (based on these bootstrapped samples), so this phenomenon is thus not an artefact of the way the local Gaussian spectral densities are estimated.

**The frequency-component:** Figure F.2 indicates that the block length sensitivity, as measured by $D(f_v^m(\omega))$, for the adjusted resampling strategy from theorem F.4 is rather small. But does this imply that these block lengths should be considered *equally good* or *equally bad*? That can not be concluded from fig. F.2 alone, and it is thus necessary to also consider a plot that takes the frequency-dimension into account, as is done in fig. F.3 for the four block lengths $L \in \{10, 25, 50, 69\}$.

It is clear from fig. F.3 that the differences between these estimates are rather small, and it is necessary to look closely in order to see that the pointwise confidence intervals are slightly narrower near $\omega = 0$ for the two cases $L = 25$ and $L = 50$. Moreover, the situation with minimal differences between the estimates remains unchanged even if the number of lags are increased to e.g. $h = 50$.

This might at first sight be somewhat surprising (and a source of concern), since it seems natural to assume that the block length $L$ should have a larger impact on the results. However, this result is actually quite natural to anticipate when the discussion from appendix F.4.3 is taken into account. It was there noted that the algorithm that estimates $f_v^m(\omega)$ does not use the temporal connection *between* the $(m+1)$-tuples in $\mathcal{Y}_{h:L}^\sharp = \big\{\big(Y_{t+m}^\sharp, \ldots, Y_{t+1}^\sharp, Y_t^\sharp\big)\big\}_{t=1}^{n-m}$, and that the important detail thus should be the expected number of times the different tuples would occur in $\mathcal{Y}_{h:L}^\sharp$.

It was seen in appendix F.4.3 that the majority of the tuples were expected to occur $\frac{n}{n-(L-1)}$ times, and this value hardly changes when $n = 1974$ and $L$ goes from 10 to 69.

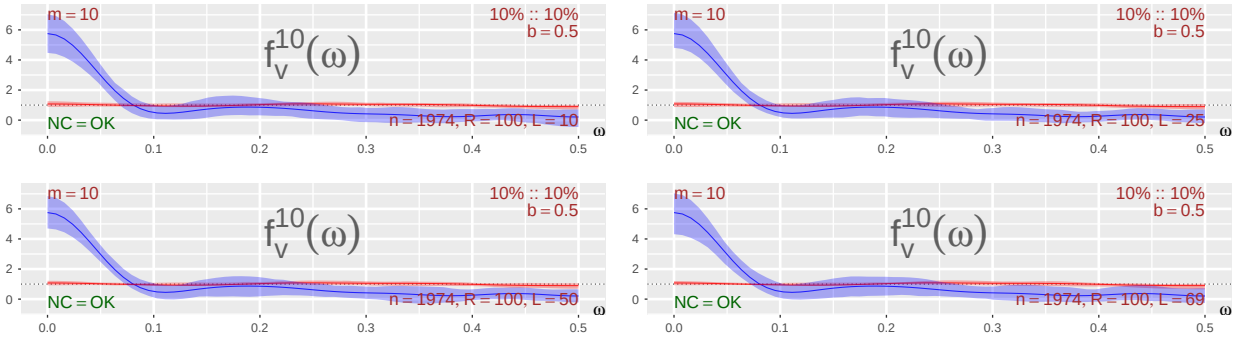


Figure F.3: Four different block lengths $L$ (from those investigated in fig. F.2) have here been used in the resampling strategy given in theorem F.4. The values of $L$ are 10, 25, 50 and 69, and this information is plotted at the lower right corner of the plots.

There are of course also differences with regard to the expected number of corrupt tuples for different block-lengths, cf. eq. (F.9) in lemma F.5, but the data in table 3 clearly indicates that this effect also can be considered minuscule.

The effect of different block lengths $L$ will of course be larger if this resampling strategy is used on a short sample, but for such samples it might not really be natural to compute the local Gaussian spectrum in the first place (since the bandwidth $\boldsymbol{b}$ in such cases must be large, and this tends to blot out local differences in the spectrum).

**An additional example:** The preceding discussion about the anticipated outcome is completely general in nature, but one might still wonder if the results in figs. F.2 and F.3 would have looked significantly differently if another case than the `dmpb`-data had been used for the investigation. This is easy to investigate since the relevant scripts trivially can be adjusted to investigate other cases too, like e.g. a single realisation from the *local trigonometric* time series, cf. figs. 7 and 8 in section 3.3.2 and the discussion in appendix G.4.

The result for the distance based box-plots for this new investigation was (as expected) very similar to the result seen in fig. F.2. The analogue of fig. F.3 is shown in fig. F.4, and it seems in fact to be the case that the differences between the pointwise confidence intervals in this case is even smaller than those observed in fig. F.3.

**Conclusion:** The preceding discussion (based on the `dmbp`-data and a *local trigonometric* example) indicates that the block length $L$ does not seem to have a major impact on the estimates and pointwise confidence intervals obtained from the adjusted resampling strategy given in theorem F.4. This simplifies the task described in section 3.4, i.e. to figure

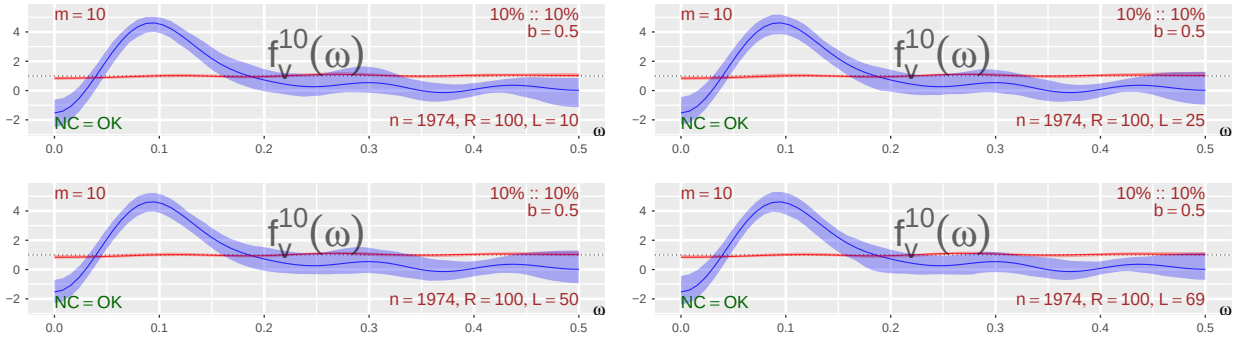


Figure F.4: A plot similar to fig. F.4, that shows that the adjusted resampling strategy from theorem F.4 also works when used on a single realisation from the local trigonometric time series used in figs. 7 and 8 in section 3.3.2, see appendix G.4 for details.

out if it for a given sample of size $n$ seems reasonable to claim that an observed difference between estimates of $f^m(\omega)$ and $f_v^m(\omega)$ is large enough to conclude that the sample at hand do have a non-Gaussian dependency structure.

For other cases, it seems natural to recommend an approach where different block lengths $L$ are tested (like seen in figs. F.3 and F.4), in order to safeguard against the possibility that the present examples for some reason both should be exceptional cases.

**Reproducibility:** The scripts needed for the reproduction of the plots in this section are included in the `R`-package `localgaussSpec` (see appendix G for further details), and the interested reader can there easily adjust the range of the block lengths to be used. It is also possible to adjust all the other tuning parameters needed for the estimation of $f_v^m(\omega)$, and it is even possible to perform the computations with the ordinary block bootstrap if so should be desired.

## F.6 What about the ordinary block bootstrap?

It was originally the ordinary block bootstrap that was used as the resampling strategy in this paper, and it could be of interest to include a few comments related to this method.

First of all, note that there are data-driven methods for the selection of the block length to be used with the block bootstrap, see e.g. Bühlmann and Künsch [1999]; Lahiri et al. [2007]; Nordman and Lahiri [2014]; Patton et al. [2009]; Politis and Romano [1994]; Politis and White [2004] — but these methods do not produce good results when used upon data with a nonlinear structure and a flat (ordinary) spectrum.

The 'problem' is easily detected from an inspection of the selection algorithms in sections 3.2 and 3.3 in Politis and White [2004], as they all have a factor $G := \sum_{h=-\infty}^{\infty} |h| R(h)$ where $R(h)$ is the lag $h$ autocovariance of the series under investigation. For a time series whose ordinary spectrum is flat, the only nonzero $R(h)$ occurs when $h = 0$, and the sum $G$ thus becomes zero in this case. This implies that the data-driven block length algorithms (both for the stationary and for the circular bootstrap) considers a very short block length to be suitable when bootstrapping the `dmbp` data — and that would obviously destroy all nonlinear structures in the data.

To the best of the authors' knowledge, it does not exist any adjustments of the algorithms used for block length selection that is suited for a situation with a flat global spectrum, and the block length $L$ thus had to be selected manually. A sensitivity analysis of the block length argument for the ordinary block bootstrap showed something similar to fig. F.2 when $L$ was large enough, e.g. the range from $L = 75$ to $L = 135$. It was mentioned at the end of section 3.4.1 in the main part that a block length of $L = 100$ had been used in an earlier draft of this paper (selected due to a visual inspection of the $\widehat{\rho}_{\boldsymbol{v}}(h)$-values seen in fig. 4, and after the testing of a few alternatives), and it can be noted that the pointwise confidence intervals then looked very similar to those based on the adjusted block bootstrap, cf. fig. 9.

The results for shorter block lengths could on the other hand be rather bad, but that is hardly surprising based on the observations in table 2 (in appendix F.3.2) about the fractions of corrupt tuples that occurs when the block bootstrap is used on a short sample.

In retrospect it is clear that the optimal resampling strategy would have been to use the block bootstrap on the derived $(m+1)$-variate time series $\left\{\left(Y_{t+m}, \ldots, Y_{t+1}, Y_t\right)\right\}_{t=1}^{n-m}$, since that would have eliminated all of the edge-effect noise — see appendix F.3.3 for the details, and an argument in favour of using the adjustment from theorem F.4 instead. Note that a resampling of the derived $(m+1)$-variate time series will have the same properties as those discussed in appendix F.4.3, and the sensitivity of the block-length argument $L$ should in this case be similar to those seen in figs. F.2 and F.3 for the adjusted resampling strategy.

The ordinary block bootstrap (working on $\{Y_t\}_{t=1}^n$) is available as a resampling strategy in the `R`-package `localgaussSpec`, but the default for this task is the adjusted block bootstrap from theorem F.4.

# Appendix G: Scripts and details related to the examples

The reproducibility of all the examples in this paper can be done by the scripts contained in the `R`-package `localgaussSpec`, and appendix G.1 explains how the interested reader

can extract these scripts. Additional details about the GARCH$(1,1)$-example seen in fig. 1, and the apARCH$(2,3)$-example seen in fig. 11, are given in appendices G.2 and G.3.

Appendix G.4 investigates the *local trigonometric* example seen in figs. 7 and 8. It starts with a theoretical investigation of the general construction of which the *local trigonometric* example is a particular realisation, and it then gives the heuristic arguments that enables this example to be used for the sanity testing of the implemented estimation algorithm.

The last part of appendix G.4 verifies that it for a large sample is possible to detect an elusive component that only occurs with probability 0.05, and it ends with some comments related to issues that can occur (under specific circumstances) when the local Gaussian machinery is used on a time series whose global spectrum does not look like white noise.

## G.1 The scripts in the `R`-package `localgaussSpec`

All the examples in this paper (and the related multivariate paper Jordanger and Tjøstheim [2017] ) can be reproduced by the scripts in the `R`-package `localgaussSpec`. This `R`-package can be installed by using '`remotes::install_github("LAJordanger/localgaussSpec")`'. The simplest way to extract the scripts from the internal storage of this `R`-package is to use the `R`-function '`LG_extract_scripts()`' after the package has been installed.

These scripts can either be used as they are (reproduction of the examples in this paper), or they can be used as templates for similar investigations of other samples/models that the user would like to investigate. In the latter case, see appendix E for some comments related to the selection of the tuning parameters of the estimation algorithm.

The reproduction of the figures requires two different scripts. The first scripts contain the code needed for the estimation of the local Gaussian autocorrelations $\rho_{\boldsymbol{v}}(h)$ for all the specified combinations of the tuning parameters, whereas the second scripts contain the code that creates the particular visualisations seen in the figures in this paper. Note that it is sufficient to use the first type of scripts in order to use the integrated `shiny`-application that enables an easy interactive investigation of the resulting estimates. The second type of scripts is first needed when one wants to put many figures into one larger grid.

## G.2 The GARCH$(1,1)$-example in fig. 1

A GARCH$(1,1)$ example was in fig. 1 used to show that the local Gaussian spectral density could detect dependency structures that the ordinary spectral density did not detect.

The following description of the standard GARCH model, introduced in Bollerslev [1986], is taken from the vignette for the `rugarch`-package Ghalanos [2020b],

$$\sigma_t^2 = \left(\omega + \sum_{j=1}^{m} \zeta_j v_{jt}\right) + \sum_{i=1}^{q} \alpha_j \varepsilon_{t-j}^2 + \sum_{i=1}^{p} \beta_j \sigma_{t-j}^2, \tag{G.1}$$

with $\sigma_t^2$ denoting the conditional variance, $\omega$ the intercept and $\varepsilon_t^2$ the residuals from the mean filtration process. The GARCH order is defined by $(q, p)$ (ARCH, GARCH), and it can also be $m$ external regressors $v_{jm}$ which are passed *pre-lagged*. Consult Ghalanos [2020a, sec. 2.2.1] for further details.

The `R` code below specifies the parameters for the GARCH$(1, 1)$ model in fig. 1.

```
library(rugarch)
.spec <- ugarchspec(
    variance.model=list(model="sGARCH",
                        garchOrder=c(1,1)),
    mean.model=list(armaOrder=c(0,0),
                    include.mean=TRUE),
    distribution.model="norm",
    fixed.pars=list(mu=0.001,
                    omega=0.00001,
                    alpha1=0.02,
                    beta1=0.95))
```

## G.3 The apARCH$(2, 3)$-example in fig. 11

The apARCH(2,3)-example seen in fig. 11 (see also figs. D.1 and D.2) had coefficients that were fitted to the `dmbp`-data by the help of the `rugarch`-package Ghalanos [2020b], and this particular model was selected after a testing procedure that tried out several thousand different variations of the GARCH-type models implemented in the `rugarch`-package.

The apARCH$(p, q)$ model (for observations $\epsilon_t$) was in Ding et al. [1993] introduced as

$$\epsilon_t = s_t e_t, \qquad e_t \sim \mathrm{N}(0, 1)\,, \tag{G.2a}$$

$$s_t^\delta = \alpha_0 + \sum_{i=1}^{p} \alpha_i \left(\left|\epsilon_{t-i}\right| - \gamma_i \epsilon_{t-i}\right)^\delta + \sum_{j=1}^{q} \beta_j s_{t-i}^\delta, \tag{G.2b}$$

where $\alpha_0 > 0$, $\delta \geq 0$, $\alpha_i \geq 0$ and $-1 < \gamma_i < 1$ for $i = 1 \ldots, p$, and $\beta_j \geq 0$ for $j = 1 \ldots, q$.

The description of this model in the `rugarch`-package, cf. Ghalanos [2020a, sec. 2.2.5], is slightly different: The constant $\alpha_0$ is there replaced with $\left(\omega + \sum_{j=1}^{m} \zeta_j v_{jt}\right)$, which is the same term that was used in eq. (G.1), see the previous section for details.

## G.4 The local trigonometric example in figs. 7 and 8

This section will discuss some topics related to the *local trigonometric* example, whose local Gaussian spectral density was investigated in figs. 7 and 8 of section 3.3.2. A few basic results related to the general construction are given in appendix G.4.1, whereas appendix G.4.2 presents the arguments that enabled this example to be used for the sanity testing of the implemented estimation algorithm.

It was noted in section 3.3.2 that the first component of the *local trigonometric* example could not be detected in the short sample investigated in that section, but it is possible to detected it when the sample-size is large enough, cf. fig. G.3 in appendix G.4.3.

Finally, appendix G.4.4 highlights issues that can occur (under specific circumstances) when this machinery is used on a time series whose global spectrum does not look like white noise, as seen in fig. G.4 where the $m$-truncated local Gaussian spectrum has been estimated for samples from a deterministic function perturbed by very low random fluctuations.

### G.4.1 Some properties of the general construction

Recall that the *local trigonometric* example is a particular case of a general construction, in which a new time series $\{Y_t\}_{t\in\mathbb{Z}}$ is constructed by the following scheme:

1. Select $r$ time series $\{C_i(t)\}_{i=1}^{r}$.
2. Select a random variable $I$ with values in the set $\{1, \ldots, r\}$, and use this to sample a collection of indices $\{I_t\}_{t\in\mathbb{Z}}$ (i.e. for each $t$ an independent realisation of $I$ is taken). Let $p_i := \mathrm{P}(I_i = i)$ denote the probabilities for the different outcomes.
3. Define $Y_t$ by means of the equation

$$Y_t := \sum_{i=1}^{r} \mathbb{1}\{I_t = i\} \cdot C_i(t). \tag{G.3}$$

The basic properties of $\{Y_t\}_{t\in\mathbb{Z}}$ can be expressed relatively those of $\{C_i(t)\}_{i=1}^{r}$, as seen in the following result.

**Lemma G.1.** *With $\{Y_t\}_{t\in\mathbb{Z}}$ as defined above, it follows that:*

*(a)* $\mathrm{E}[Y_t] = \sum_{i=1}^{r} p_i \cdot \mathrm{E}[C_i(t)]$

*(b)* $\mathrm{E}\big[Y_{t+h} \cdot Y_t\big] = \begin{cases} \sum_{i=1}^{r}\sum_{j=1}^{r} p_i \cdot p_j \cdot \mathrm{E}[C_i(t+h) \cdot C_i(t)] & h \neq 0 \\ \sum_{i=1}^{r} p_i \cdot \mathrm{E}[C_i(t)^2] & h = 0 \end{cases}$

(c) $\mathrm{Cov}\big(Y_{t+h}, Y_t\big) = \begin{cases} \sum_{i=1}^r \sum_{j=1}^r p_i \cdot p_j \cdot \mathrm{Cov}\big(C_i(t+h), C_j(t)\big) & h \neq 0 \\ \sum_{i=1}^r p_i \cdot \mathrm{E}[C_i(t)^2] - \big(\sum_{i=1}^r p_i \cdot \mathrm{E}[C_i(t)]\big)^2 & h = 0 \end{cases}$

(d) *The additional assumption that* $C_i(t)$ *and* $C_j(t)$ *are independent when* $i \neq j$*, simplifies the* $h \neq 0$ *case to:* $\mathrm{Cov}\big(Y_{t+h}, Y_t\big) = \sum_{i=1}^r p_i^2 \cdot \mathrm{Cov}(C_i(t+h), C_i(t))$.

*Proof.* The random variable $I_t$ that produces the set of indices is independent of $C_i(t)$, and item (a) thus follows without further ado. For the $h \neq 0$ case of item (b) it is sufficient to note that $I_{t+h}$ and $I_t$ then are independent, and it follows that $\mathrm{E}\big[\mathbb{1}\{I_{t+h} = i\} \cdot \mathbb{1}\{I_t = j\}\big] = \mathrm{E}\big[\mathbb{1}\{I_{t+h} = i\}\big] \cdot \mathrm{E}[\mathbb{1}\{I_t = j\}] = \mathrm{P}\big(I_{t+h} = i\big) \cdot \mathrm{P}(I_t = j) = p_i \cdot p_j$. For the $h = 0$ case of item (b) it is enough to note that $\mathbb{1}\{I_t = i\} \cdot \mathbb{1}\{I_t = j\} = 0$ when $i \neq j$, which together with $\mathbb{1}\{I_t = i\} \cdot \mathbb{1}\{I_t = i\} = \mathbb{1}\{I_t = i\}$ gives the required expression. The statements in items (c) and (d) follows trivially from those in items (a) and (b). □

The key idea in the local trigonometric example is that the $r$ time series $C_i(t)$ all should be 'cosines with some noise', since this implies (given a reasonable parameter configuration) that it should be possible to present a decent guesstimate with regard to the expected shape of the $m$-truncated local Gaussian spectral density (for some carefully selected tuning parameters of the estimation algorithm). The global spectrum in this case will not be flat, but it will for low truncation levels be 'flat enough' for the purpose of showing that the global spectrum does not detect the underlying frequencies whereas the local Gaussian spectral density function can do that task.

The following result reiterates the $C_i(t)$-definition used in the local trigonometric example, and it presents some basic properties related to this definition.

**Lemma G.2.** *Let* $C_i(t) = L_i + A_i(t) \cdot \cos(2\pi\alpha_i t + \varphi_i)$*, be defined in the following manner:* $L_i$ *and* $\alpha_i$ *are constants that respectively defines the horizontal base-line and the frequency. The amplitude* $A_i(t)$ *are for each* $t$ *uniformly distributed on an interval* $[a_i, b_i]$*, and* $A_i(t+h)$ *and* $A_i(t)$ *are independent when* $h \neq 0$*. The phase-adjustment* $\varphi_i$ *are uniformly drawn (one time for each realisation) from the interval between* $0$ *and* $2\pi$*, and it is moreover assumed that the stochastic processes* $\varphi_i$ *and* $A_i(t)$ *are independent of each other.*

(a) $\mathrm{E}[C_i(t)] = L_i$

(b) $\mathrm{E}[C_i(t+h) \cdot C_i(t)] = \begin{cases} L_i^2 + \frac{\pi}{4} \cdot (a_i^2 + 2a_i b_i + b_i^2) \cdot \cos(2\pi\alpha_i \cdot h) & h \neq 0 \\ L_i^2 + \frac{\pi}{3} \cdot (a_i^2 + a_i b_i + b_i^2) & h = 0 \end{cases}$

(c) $\mathrm{Cov}(C_i(t+h), C_i(t)) = \begin{cases} \frac{\pi}{4} \cdot (a_i^2 + 2a_i b_i + b_i^2) \cdot \cos(2\pi\alpha_i \cdot h) & h \neq 0 \\ \frac{\pi}{3} \cdot (a_i^2 + a_i b_i + b_i^2) & h = 0 \end{cases}$

*Proof.* This is a consequence of the independence of the two stochastic processes $A_i(t)$ and $\varphi_i$, and the basic observations: $\mathrm{E}[A_i(t)] = \frac{1}{2} \cdot (a_i + b_i)$, $\mathrm{E}[A_i^2(t)] = \frac{1}{3} \cdot (a_i^2 + a_i b_i + b_i^2)$, $\mathrm{E}[\cos(2\pi\alpha_i t + \varphi_i)] = 0$ and $\mathrm{E}[\cos(2\pi\alpha_i(t+h) + \varphi_i) \cdot \cos(2\pi\alpha_i t + \varphi_i)] = \pi \cdot \cos(2\pi\alpha_i \cdot h)$. The proof of item (a) is trivial. For item (b) it suffices to observe that the $h \neq 0$ case contains $\mathrm{E}[A_i(t)]^2$ as a factor, whereas the $h = 0$ case contains $\mathrm{E}[A_i(t)^2]$ as a factor. Item (c) follows from items (a) and (b). □

Finally, the *local trigonometric* example is obtained by using $r$ time series $C_i(t)$, of the form given in lemma G.2, in the construction of the time series $Y_t$, i.e.

$$Y_t = \sum_{i=1}^{r} \mathbb{1}\{I_t = i\} \cdot (L_i + A_i(t) \cdot \cos(2\pi\alpha_i t + \varphi_i)), \tag{G.4}$$

where it furthermore is assumed that the $i$-indexed stochastic variables $A_i(t)$ and $\varphi_i$ are independent of the $j$-indexed variants when $i \neq j$. It now follows from lemmas G.1 and G.2 that the $h \neq 0$ correlation of the time series $Y_t$ in eq. (G.4) is given by

$$\rho_Y(h) = \frac{\frac{\pi}{4} \cdot \sum_{i=1}^{r} p_i^2 \cdot (a_i^2 + 2a_i b_i + b_i^2) \cdot \cos(2\pi\alpha_i \cdot h)}{\sum_{i=1}^{r} p_i \cdot \left[L_i^2 + \frac{\pi}{3} \cdot (a_i^2 + a_i b_i + b_i^2)\right] - \left(\sum_{i=1}^{r} p_i \cdot L_i\right)^2}. \tag{G.5}$$

An inspection of eq. (G.5) reveals that it is fairly easy to find a parameter configuration for which the numerator is rather small compared to the denominator. This is of course not white noise, but the key idea is that it is close enough to white noise to make it impossible to deduce anything about the underlying frequencies $\alpha_i$ based on the ordinary spectrum.

### G.4.2 The heuristic argument that motivates the local trigonometric example

This section starts with an outline that shows how it is possible to select the parameters of the *local trigonometric* time series from eq. (G.4) in such a manner that some specified key features should be present after the pseudo-normalisation of a sample. It is with regard to this also necessary to take into account the tuning-parameters of the estimation algorithm (i.e. the bandwidth $\boldsymbol{b}$), since these must be adjusted relative to the size $n$ of the sample. The last part of this section considers the example used in section 3.3.2, and the discussion related to fig. G.1 will pinpoint why this hand-waving approach actually works.

**The heuristic argument:** The amplitude $A_i(t)$ is uniformly distributed on $[a_i, b_i]$, and it thus follows that all the observations from the $C_i$ component lies in the interval $\mathcal{I}_i = [L_i - b_i, L_i + b_i]$. The first key requirement is that the $r$ intervals $\mathcal{I}_i$ should have a minimal amount of overlap, and it is moreover for simplicity natural to require that the base-lines are ordered as follows $L_1 < L_2 < \cdots < L_r$.

The base lines $L_i$ do occur in the denominator of $\rho_Y(h)$, cf. eq. (G.5), but for the purpose of the local Gaussian spectral density investigation it is the corresponding values after the pseudo-normalisation that is of interest. This implies that the values of $L_i$, $a_i$ and $b_i$ are somewhat irrelevant, since minor modifications of them will return exactly the same pseudo-normalised sample.

The key ingredient with regard to the pseudo-normalised version of the sampled values that lies in a given interval $\mathcal{I}_i$, is the the specification of the probability $p_i := \mathrm{P}(I_t = i)$. Assuming that the intervals $\mathcal{I}_i$ does not overlap, it is clear that it for a sample of size $n$ will be natural to assume that approximately $p_i \cdot n$ of the observations should lie in the interval $\mathcal{I}_i$ — and the symmetry of the cosine around its baseline then implies that approximately one half of these $p_i \cdot n$ observations should lie below $L_i$ and the other half above it.

It follows from this that the base-line $L_1$ of the $C_1(t)$ component should occur near the $v_1 := \frac{1}{2}p_1$ percentile of the sample, the 'border-line' between $C_1(t)$ and $C_2(t)$ near the $p_1$ percentile, the base-line $L_2$ of the $C_2(t)$ near the $v_2 := \left(p_1 + \frac{1}{2}p_2\right)$ percentile, and so on. This implies that the part of the sample that lies in the interval $I_i$ should correspond to the observations between the two percentiles $v_i - \frac{1}{2}p_i$ and $v_i + \frac{1}{2}p_i$, and this part should moreover look like a random selection of $p_i \cdot n$ observations from the $C_i(t)$ component.

The idea now is that the pseudo-normalisation of the $I_i$-part of the sample still should contain a structure that reveals the frequency $\alpha_i$ of the underlying cosine, and that it thus (given a suitable combination of point $\boldsymbol{v}$ and bandwidth $\boldsymbol{b}$) should be possible to get a result that looks approximately like the result obtained when $f^m_{\boldsymbol{v}}(\omega)$ is estimated for a single cosine with a frequency equal to $\alpha_i$.

It is possible to select the probabilities $\{p_i\}_{i=1}^r$ such that the percentile $v_i$ of the base-line $L_i$ corresponds directly to the diagonal point $\boldsymbol{v}$ for which $f^m_{\boldsymbol{v}}(\omega)$ should be estimated, but this does not take into account that the pseudo-normalised version of the $I_i$-part of the sample in general will not be symmetric around $\Phi^{-1}(v_i)$ (with one exception when $v_i = 0.5$). The probabilities $\{p_i\}_{i=1}^r$ should be selected such that the $v_i$ percentile lies closer to the center than the percentile corresponding to $\boldsymbol{v}$.

Given a configuration of probabilities $\{p_i\}_{i=1}^r$, and furthermore assuming that the probability $p_i$ for the $C_i(t)$ component is sufficiently large relatively the sample size $n$, it will now be possible to find a point $\boldsymbol{v}$ and a bandwidth $\boldsymbol{b}$ such that the estimate of $f^m_{\boldsymbol{v}}(\omega)$ has the predicted shape with a peak at the frequency $\alpha_i$ that is used in $C_i(t)$

It is possible to construct *local trigonometric* examples where some of the components $C_i(t)$ are impossible to detect for a given sample size $n$, but they could still be detected when a larger sample is used, cf. the example discussed in appendix G.4.3.

**The case investigated in section 3.3.2:** The heuristic arguments outlined above were used in order to create the *local trigonometric* example in section 3.3.2. The initial requirements for the construction of that particular example were that the sample should have the same length as the `dmbp`-data, i.e. $n = 1974$, that the bandwidth should be 0.5, and that the investigation should be performed at the three diagonal points corresponding to the 10%, 50% and 90% percentiles of the standard normal distribution.

The initial approach used three $C_i(t)$-components with equal probability of being selected. An additional new first component $C_1(t)$ was then added, with $p_1 = 0.05$ (and with a corresponding reduction of the next probability to $p_2 = \frac{1}{3} - 0.05$). This adjustment was done in order to get more mileage out of the example, since it then also could be used to highlight that the local Gaussian spectral density in some cases might not have enough observations available to detect all the local features. Note that the elusive first component can be detected when the sample size increases, cf. appendix G.4.3 for details.

The explicit expression of the *local trigonometric* time series used in section 3.3.2 is given by the following equation,

$$Y_t \coloneqq \sum_{i=1}^{4} \mathbb{1}\{I_t = i\} \cdot (L_i + A_i(t) \cdot \cos(2\pi\alpha_i t + \varphi_i)) \,, \tag{G.6}$$

where the probabilities $p_i \coloneqq \mathrm{P}(I_t = i)$ are given by $(0.05, 1/3 - 0.05, 1/3, 1/3)$, and the frequencies $\alpha_i$ are given by $(0.267, 0.091, 0.431, 0.270)$. The base-lines $L_i$ are given by the values $(-2, -1, 0, 1)$, and the lower and upper ranges for the uniforms sampling of the amplitudes $A_i(t)$ are respectively given by $(0.5, 0.2, 0.2, 0.5)$ and $(1.0, 0.5, 0.3, 0.6)$. Recall that these latter values are not really of interest with regard to the pseudo-normalised version of the sample, and the only requirement regarding these should be that they are selected in order to give a minimal amount of overlap between the different components. The phase-adjustments $\varphi_i$ are uniformly selected from the interval $[0, 2\pi)$, one time for each realisation of a sample from $Y_t$.

Figure G.1 shows a *simplified* excerpt of length 100 from one realisation of $Y_t$. The amplitudes $A_i(t)$ have here for the simplicity of the present discussion been fixed to the values $(1.0, 0.5, 0.3, 0.5)$ since it is of importance to emphasise which one of the underlying 'hidden' components $C_i(t)$ (shown as dotted curves in the top panel) that was selected in this case (the phase-adjustments $\varphi_i$ in this particular realisation are $(0.52, 2.57, 3.24, 2.49)$). The center panel of fig. G.1 shows an estimate of the $m$-truncated (global) spectral density $f^m(\omega)$, based on 100 independent samples of length 1974 and with a 90% pointwise confidence interval that shows that it is viable to claim that this particular process be-

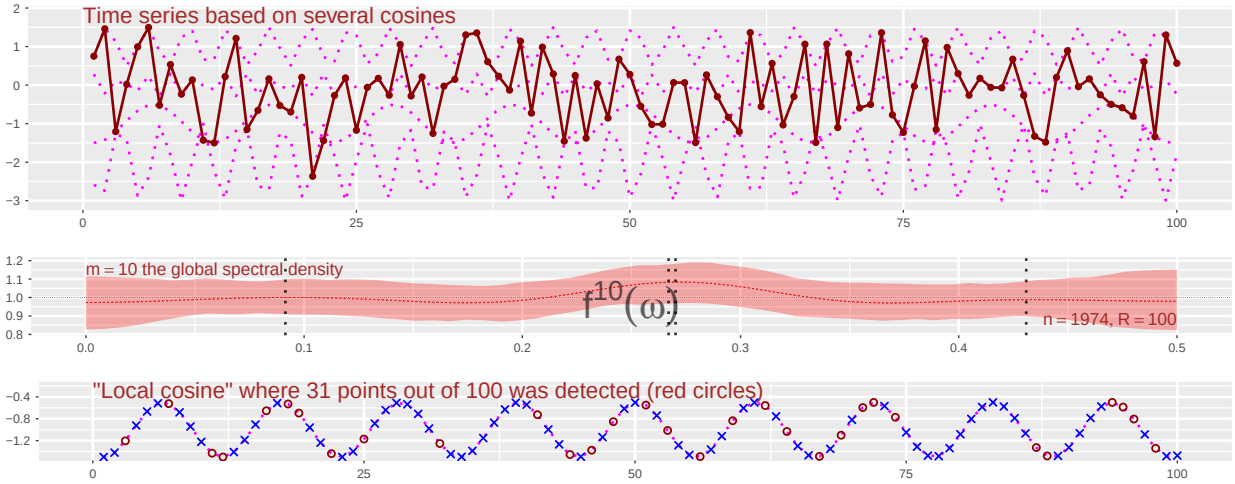


Figure G.1: Top: Short excerpt from artifical example based on *hidden trigonometric components*. Center: Estimated (truncated) global spectral density (*hidden frequencies* indicated with vertical lines). Bottom: Local cosine showing the detected points at the local level centered at -1. Further details in the main text.

haves almost like white noise. Note that the vertical lines in the center panel shows the frequencies $\alpha_i$ that was used in eq. (3.2).

The bottom panel of fig. G.1 is the one of major interest for the present discussion, i.e. it is the one from which it is possible to provide an explanation for the expected shape of the local Gaussian spectral density, at some particularly designated points $\boldsymbol{v}$ (given a suitable bandwidth $\boldsymbol{b}$). First of all, the bottom panel shows *one* of the cosines from the top panel, the circles represents the points from the top panel that happened to lie on this particular cosine — and the crosses represents all the remaining points (at integer valued times $t$) of the cosine. Recall that these points are from the simplified realisation where the amplitudes $A_i(t)$ are constant, and that the actual values thus would be distorted a bit from those observed here.

The circles can be considered as a randomly selected collection of points from a time series like the one investigated in fig. 6 (single cosine function with some white noise), and the main point of interest is that it (for a sufficiently long time series, and a sufficiently large bandwidth $\boldsymbol{b}$) will be the case that the estimated local Gaussian autocorrelations based on this scarce subset might be quite close to the estimates obtained if all the points had been available. The rationale for this claim is related to the way that the local Gaussian auto-correlation at lag $h$ (at a given point $\boldsymbol{v}$) is computed from the sets of bivariate points

$\big(Y_{t+h}, Y_t\big)$. In particular: It might not have a detrimental effect on the resulting estimate if some of these lag $h$ pairs are removed at random, as long as the remaining number of pairs is large enough. Based on this idea, it can thus be argued that the local Gaussian spectral density estimated from the collection of the circled-marked points should be fairly close to the situation shown in fig. 6, at least if the time series under investigation is sufficiently long.

This final heuristic graphical argument is the reason for the guesstimate that the $m$-truncated local Gaussian spectral densities (for the three points $\boldsymbol{v}$ corresponding to the 10%, 50% and 90% percentiles) should have an overall shape that resembles the one seen for the single cosine example seen in fig. 6.

It did turn out, cf. figs. 7 and 8, that the guesstimate based on these heuristic arguments in fact did hold true, and the $m$-truncated local Gaussian spectral densities did in fact detect the specified frequencies $\alpha_i$ in $(0.091, 0.431, 0.270)$ at the three targeted points $\boldsymbol{v}$.

Note that the frequency 0.267 corresponding to the $C_1(t)$ component could not be detect based on only $n = 1974$ observations, since the probability $p_1 = 0.05$ requires an investigation far out in the lower tail. It is however possible to detect it whit a much larger sample size, cf. the discussion in the next section.

### G.4.3 Detecting the $C_1(t)$ component of the local trigonometric example

The *local trigonometric* example seen in figs. 7 and 8 of section 3.3.2, cf. eq. (G.6) for the definition, contains a component $C_1(t)$ that goes undetected when the sample size of $n = 1974$ is used. The reason for the elusiveness of the $C_1(t)$ component is that it has a probability of $p_1 = 0.05$ of being selected, which implies that it is expected to only find 98.7 observations from this component when $n = 1974$.

The 'border' between the observations from the $C_1(t)$ and $C_2(t)$ components should occur near the 5% percentile, but it is necessary to 'zoom in' on a point $\boldsymbol{v}$ that lies farther out in the tail than $p_1/2 = 0.025$. This requirement occurs since the estimate of $f^m_{\boldsymbol{v}}(\omega)$ should avoid 'contamination' from the observations from the $C_2(t)$ component.

Based on the idea that it might be necessary to go all the way out to the 1%, it seemed natural to attempt an investigation based on $n = 25000$ observations. Since the point $\boldsymbol{v}$ now is far out in the lower tail, e.g. the 1% percentile of the standard normal distribution is $-2.326$, it seemed reasonable to use the bandwidth $\boldsymbol{b} = (0.4, 0.4)$.

The heatmap and distance plots in fig. G.2 is based on an investigating of a single realisation, that included percentiles based on values starting from 2 bandwidths below the 5% percentile and ending at 1/2 bandwidth below the 5% percentile, i.e. the diagonal points starts at approximately the 0.72% percentile and ends at the 3.25% percentile.

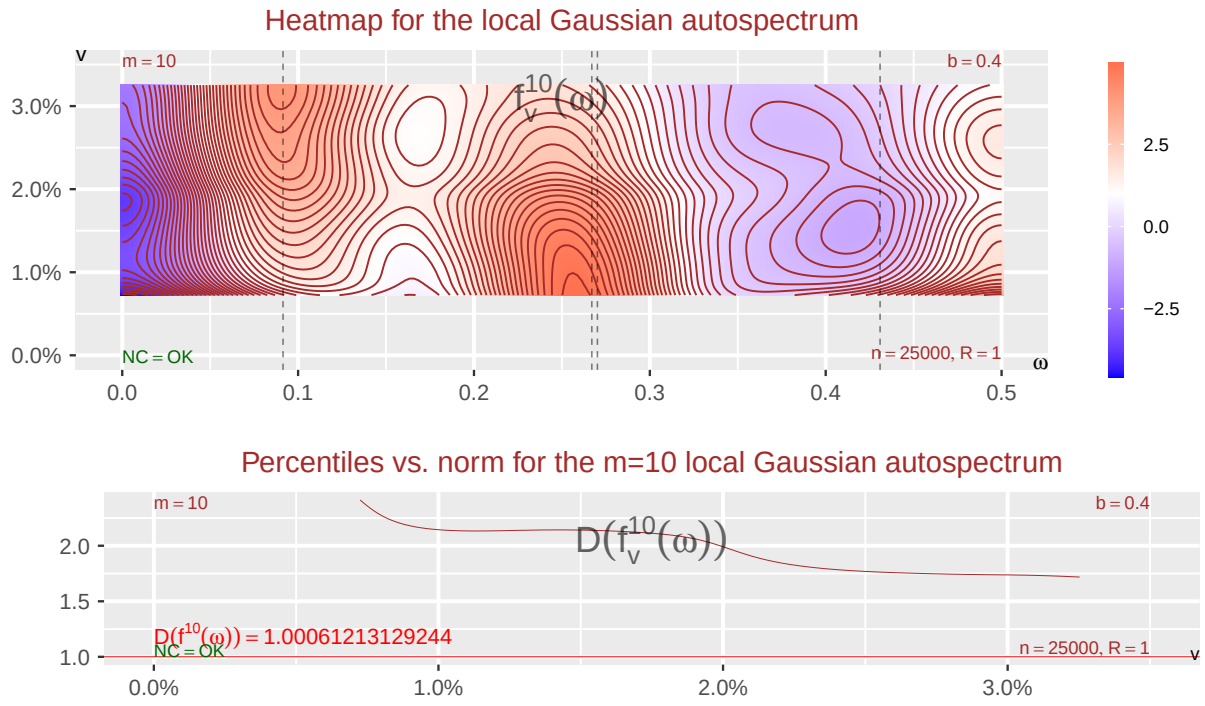


Figure G.2: A heatmap+distance plot used to search for an 'optimal' percentile that can reveal the $\alpha_1$ frequency in the lower tail of the *local trigonometric*.

It is no surprise that the $C_2(t)$ component completely dominates at the 3.25% percentile, and it can be seen that it is necessary to go down to at least the 1% percentile in order to detect a peak close to the frequency $\alpha_1 = 0.267$ of the $C_1(t)$ component. Note that fig. G.2 is based on only 1 single realisation, and other realisations might look slightly different.

Figure G.3 shows the situation when $R = 100$ replicates are used to estimate $f_{\boldsymbol{v}}^m(\omega)$ at the diagonal point $\boldsymbol{v}$ that corresponds to the 1% percentile. This shows that $\widehat{f}_{\boldsymbol{v}}^m(\omega)$ in this case has the expected 'cosine'-shape, and the peak is at the frequency $\alpha_1 = 0.267$ of the $C_1(t)$ component.

### G.4.4 Beware of deterministic global structures under small noise

It was seen in the *local trigonometric* example investigated in the preceding sections, that a peak of $\widehat{f}_{\boldsymbol{v}}^m(\omega)$, like the one seen in fig. G.2, corresponded to a frequency $\alpha$ of some underlying cosine-function (that $\widehat{f}^m(\omega)$ did not detect). It should here be emphasised that

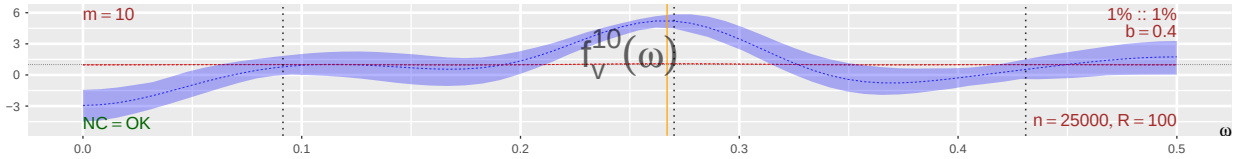


Figure G.3: The detection of the $\alpha_1$ frequency in the lower tail of the *local trigonometric* example requires a large sample and an investigation far out in the lower tail.

the *local trigonometric* example was fine tuned in order to test the sanity of the implemented estimation algorithm — and it would thus be a logical fallacy to conclude that a similar peak of $\widehat{f}_v^m(\omega)$ (not present in $\widehat{f}^m(\omega)$) always could be interpreted in the same manner for general non-Gaussian time series.

An investigation of this issue can be seen in fig. G.4, where an extreme version of the case investigated in fig. 6 are presented. The setup is similar to the one from fig. 6, i.e. the plots are based on 100 samples of length 1974 from a model of the form $Y_t = \cos(2\pi\alpha t + \varphi) + w_t$, where $\alpha = 0.302$ (as before), but the standard deviation of the Gaussian white noise $w_t$ has now been reduced to $\sigma = 0.05$.

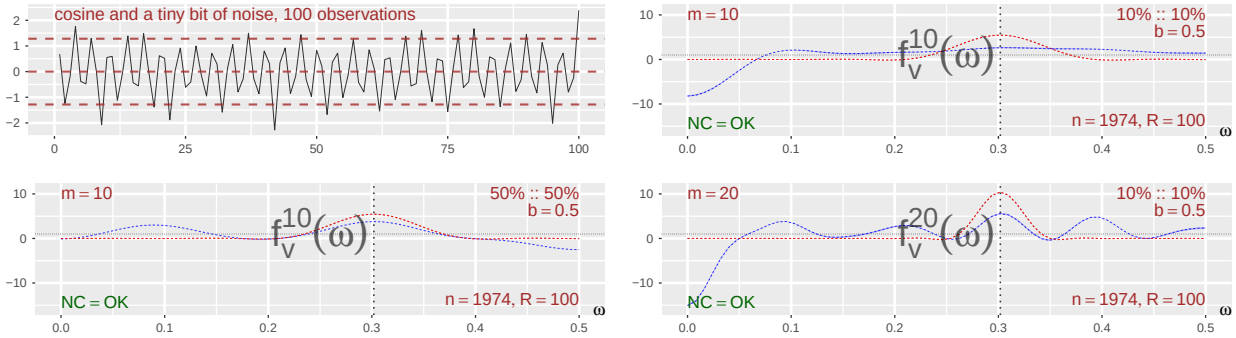


Figure G.4: Pseudo-normalised single cosine with small noise.

The low value of the standard deviation $\sigma$ implies that samples from this model have a very clear periodic behaviour, as can be seen from the plots in fig. G.4, where the 90% confidence intervals are almost indistinguishable from the mean of the estimates. This clear periodicity is also evident from the trace shown in the upper left panel of fig. G.4, where the 100 first pseudo-normalised observations of one of the samples are presented.

Figure G.4 contains estimates of the local and global spectra, with focus on the points in the lower tail and the center for the truncation level $m = 10$, and for the lower tail when

$m = 20$. The additional peaks seen at the center is due to the kernel function $K$ that is used in the estimation algorithm — in particular $K$ works on the $h$-lagged pairs $\left(Y_{t+h}, Y_t\right)$, the contribution becomes negligible for pairs far away from $\boldsymbol{v}$, the pairs that do contribute give the impression that several 'local frequencies' might be present, but the underlying model has only one single frequency.

The case in fig. G.4 is extreme since the noise-term is minuscule. Because the local Gaussian correlation is based on a continuous distribution assumption and the use of a kernel function, similar difficulties can be expected for other deterministic functions embedded in low noise. One possible way out might be to consider an approach where a parametric model is fitted first to the data and then examine the residuals with a global and a local spectral analysis.

# References


Basseville, M., 2013. Divergence measures for statistical data processing — An annotated bibliography. Signal Processing 93, 621 – 633. URL: http://www.sciencedirect.com/science/article/pii/S0165168412003222, doi:10.1016/j.sigpro.2012.09.003.

Berentsen, G.D., Tjøstheim, D., 2014. Recognizing and visualizing departures from independence in bivariate data using local Gaussian correlation. Statistics and Computing 24, 785–801. URL: http://dx.doi.org/10.1007/s11222-013-9402-8, doi:10.1007/s11222-013-9402-8.

Billingsley, P., 2012. Probability and Measure. Aniversary ed., Wiley.

Birr, S., Kley, T., Volgushev, S., 2019. Model assessment for time series dynamics using copula spectral densities: A graphical tool. Journal of Multivariate Analysis 172, 122 – 146. URL: http://www.sciencedirect.com/science/article/pii/S0047259X18301842, doi:10.1016/j.jmva.2019.03.003. Dependence Models.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307 – 327. URL: http://www.sciencedirect.com/science/article/pii/0304407686900631, doi:10.1016/0304-4076(86)90063-1.

Brockwell, P.J., Davis, R.A., 1986. Time Series: Theory and Methods. Springer-Verlag New York, Inc., New York, NY, USA.

Bühlmann, P., Künsch, H.R., 1999. Block length selection in the bootstrap for time series. Computational Statistics & Data Analysis 31, 295–310. URL: http://www.sciencedirect.com/science/article/pii/S0167947399000146, doi:10.1016/S0167-9473(99)00014-6.

Burman, P., Chow, E., Nolan, D., 1994. A Cross-Validatory Method for Dependent Data. Biometrika 81, 351–358. URL: http://www.jstor.org/stable/2336965.

Chen, T., Sun, Y., Li, T.H., 2019. A Semi-Parametric Estimation Method for the Quantile Spectrum with an Application to Earthquake Classification Using Convolutional Neural Network. arXiv:1910.07155.

Davydov, Y.A., 1968. Convergence of Distributions Generated by Stationary Stochastic Processes. Theory of Probability and Application 13, 691–696. URL: http://dx.doi.org/10.1137/1113086, doi:10.1137/1113086.

Ding, Z., Granger, C.W., Engle, R.F., 1993. A long memory property of stock market returns and a new model. Journal of Empirical Finance 1, 83–106. URL: http://www.sciencedirect.com/science/article/pii/092753989390006D, doi:10.1016/0927-5398(93)90006-D.

Efron, B., 1979. Bootstrap methods: Another look at the jackknife. Ann. Statist. 7, 1–26. URL: https://doi.org/10.1214/aos/1176344552, doi:10.1214/aos/1176344552.

Fan, J., Yao, Q., 2003. Nonlinear Time Series: Nonparametric and Parametric Methods. Springer.

Georgiou, T.T., 2007. Distances and Riemannian Metrics for Spectral Density Functions. IEEE Transactions on Signal Processing 55, 3995–4003. doi:10.1109/TSP.2007.896119.

Ghalanos, A., 2020a. Introduction to the rugarch package (Version 1.4-2). URL: https://cran.r-project.org/web/packages/rugarch/vignettes/Introduction_to_the_rugarch_package.pdf.

Ghalanos, A., 2020b. rugarch: Univariate GARCH models. URL: https://cran.r-project.org/package=rugarch. r package version 1.4-2.

Hjort, N.L., Jones, M.C., 1996. Locally parametric nonparametric density estimation. Ann. Statist. 24, 1619–1647. URL: http://dx.doi.org/10.1214/aos/1032298288, doi:10.1214/aos/1032298288.

Horn, R.A., Johnson, C.R., 2012. Matrix Analysis. 2nd ed., Cambridge University Press, New York, NY, USA.

Jordanger, L.A., Tjøstheim, D., 2017. Nonlinear cross-spectrum analysis via the local gaussian correlation. arXiv preprint arXiv:1708.02495 URL: https://arxiv.org/abs/1708.02495, arXiv:https://arxiv.org/abs/1708.02495.

Klimko, L.A., Nelson, P.I., 1978. On Conditional Least Squares Estimation for Stochastic Processes. Ann. Statist. 6, 629–642. URL: http://dx.doi.org/10.1214/aos/1176344207, doi:10.1214/aos/1176344207.

Kullback, S., Leibler, R.A., 1951. On information and sufficiency. Ann. Math. Statist. 22, 79–86. URL: http://dx.doi.org/10.1214/aoms/1177729694, doi:10.1214/aoms/1177729694.

Künsch, H.R., 1989. The Jackknife and the Bootstrap for General Stationary Observations. The Annals of Statistics 17, 1217–1241. URL: http://www.jstor.org/stable/2241719, doi:10.1214/aos/1176347265.

Lahiri, S.N., Furukawa, K., Lee, Y.D., 2007. A nonparametric plug-in rule for selecting optimal block lengths for block bootstrap methods. Statistical Methodology 4, 292–321. URL: http://www.sciencedirect.com/science/article/pii/S1572312706000505, doi:10.1016/j.stamet.2006.08.002.

Masry, E., Tjøstheim, D., 1995. Nonparametric Estimation and Identification of Nonlinear ARCH Time Series Strong Convergence and Asymptotic Normality: Strong Convergence and Asymptotic Normality. Econometric Theory 11, 258–289. URL: http://EconPapers.repec.org/RePEc:cup:etheor:v:11:y:1995:i:02:p:258-289_00.

Nordman, D.J., Lahiri, S.N., 2014. Convergence rates of empirical block length selectors for block bootstrap. Bernoulli 20, 958–978. URL: http://dx.doi.org/10.3150/13-BEJ511, doi:10.3150/13-BEJ511.

Otneim, H., Tjøstheim, D., 2017. The locally Gaussian density estimator for multivariate data. Statistics and Computing 27, 1595–1616. URL: https://doi.org/10.1007/s11222-016-9706-6, doi:10.1007/s11222-016-9706-6.

Patton, A., Politis, D.N., White, H., 2009. Correction to "Automatic Block-Length Selection for the Dependent Bootstrap" by D. Politis and H. White. Econometric Reviews 28, 372–375. URL: http://dx.doi.org/10.1080/07474930802459016, doi:10.1080/07474930802459016, arXiv:http://dx.doi.org/10.1080/07474930802459016.

Politis, D.N., Romano, J.P., 1992. A General Resampling Scheme for Triangular Arrays of $\alpha$-Mixing Random Variables with Application to the Problem of Spectral Density Estimation. The Annals of Statistics 20, 1985–2007. URL: https://www.jstor.org/stable/2242377, doi:10.1214/aos/1176348899.

Politis, D.N., Romano, J.P., 1994. Limit theorems for weakly dependent Hilbert space valued random variables with application to the stationary bootstrap. Statistica Sinica 4, 461–476. URL: http://www.jstor.org/stable/24305527.

Politis, D.N., White, H., 2004. Automatic Block-Length Selection for the Dependent Bootstrap. Econometric Reviews 23, 53–70. URL:

http://dx.doi.org/10.1081/ETC-120028836, doi:10.1081/ETC-120028836, arXiv:http://dx.doi.org/10.1081/ETC-120028836.

Racine, J., 2000. Consistent cross-validatory model-selection for dependent data: *hv*-block cross-validation. Journal of Econometrics 99, 39–61. URL: http://www.sciencedirect.com/science/article/pii/S0304407600000300, doi:10.1016/S0304-4076(00)00030-0.

Shao, J., 1993. Linear Model Selection by Cross-validation. Journal of the American statistical Association 88, 486–494. URL: http://www.tandfonline.com/doi/abs/10.1080/01621459.1993.10476299, doi:10.1080/01621459.1993.10476299, arXiv:http://www.tandfonline.com/doi/pdf/10.1080/01621459.1993.10476299.

Taniguchi, M., Kakizawa, Y., 2000. Asymptotic Theory of Statistical Inference for Time Series. Springer.

Tjøstheim, D., Hufthammer, K.O., 2013. Local Gaussian correlation: A new measure of dependence. Journal of Econometrics 172, 33 – 48. URL: http://www.sciencedirect.com/science/article/pii/S0304407612001741, doi:10.1016/j.jeconom.2012.08.001.

Volkonskii, V.A., Rozanov, Y.A., 1959. Some Limit Theorems for Random Functions. I. Theory of Probability and Application 4, 178–197. URL: http://epubs.siam.org/doi/10.1137/1104015, doi:10.1137/1104015.