

Nonlinear spectral analysis via the local Gaussian correlation

Lars Arne Jordanger Dag Tjøstheim

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* $\rho_v(h)$ introduced in Tjøstheim and Hufthammer (2013), and these *local measures of dependence* can be used to construct the *local Gaussian spectral density* $f_v(\omega)$ that is presented in this paper. A key feature of $f_v(\omega)$ is that it coincides with $f(\omega)$ for Gaussian time series, which implies that $f_v(\omega)$ 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 $f_v(\omega)$ can indicate nonlinear traits, which potentially might discover *local periodic phenomena* that goes undetected in an ordinary spectral analysis.

1 Introduction

It is well known that stock returns behave in an asymmetric manner, i.e. that they, as noted in e.g. Hong et al. (2007), ‘more often move with the market when the market goes down than when it goes up’. An asymmetry in (Y_{t+h}, Y_t) can not be detected by the autocorrelation $\gamma(h)$, which renders the corresponding spectral density $f(\omega)$ an inadequate tool for this kind of phenomenon. Several generalisations of $f(\omega)$ have been developed based on the idea that the second order moment $\gamma(h)$ could be replaced with some other measure of dependence (to be described later on), and this paper uses this approach to define a *local Gaussian spectral density* $f_v(\omega)$ based on the *local Gaussian correlations* $\rho_v(h)$ from Tjøstheim and Hufthammer (2013).

If a weakly stationary time series satisfies the additional requirement that the autocovariances are absolutely summable, then the spectral density $f(\omega)$ is the Fourier transform of $\{\gamma(h)\}_{h \in \mathbb{Z}}$, i.e.

$$f(\omega) := \sum_{h \in \mathbb{Z}} \gamma(h) \cdot e^{-2\pi i \omega h}. \quad (1.1)$$

The inverse Fourier transform gives the relation $\gamma(h) = \int_{-1/2}^{1/2} f(\omega) \cdot e^{2\pi i\omega h} d\omega$, which for $h = 0$ expresses the variance as the integral of $f(\omega)$. This enables a visual inspection of how much different frequencies contribute to the variance,¹ and peaks and troughs in the graph of $f(\omega)$ can thus reveal information about periodic properties of the time series $\{Y_t\}_{t \in \mathbb{Z}}$.

The *local Gaussian spectral density* $f_v(\omega)$ introduced in this paper will be based on a normalisation of $\{Y_t\}_{t \in \mathbb{Z}}$, which implies that the correlation $\rho(h)$ equals the covariance $\gamma(h)$, and references later on to $f(\omega)$ will thus refer to the following rescaled version,

$$f(\omega) := \sum_{h \in \mathbb{Z}} \rho(h) e^{-2\pi i\omega h}. \quad (1.2)$$

The spectral density may be an inadequate tool when the time series under investigation contains nonlinear features, like e.g. those present for GARCH-type models, where the terms of $\{Y_t\}_{t \in \mathbb{Z}}$ are uncorrelated but not independent. However, $f(\omega)$ gives a complete description of Gaussian time series, which motivates the local Gaussian approach presented in this paper, where the local Gaussian correlations $\rho_v(h)$ are used to define the *local Gaussian spectral density* $f_v(\omega)$. Note that this approach requires that $\{Y_t\}_{t \in \mathbb{Z}}$ also must be *strictly stationary*, a requirement shared with quite a few of the existing global and local extensions of $f(\omega)$, that will be briefly reviewed below.

The higher order spectra (global) generalisations of $f(\omega)$ was introduced by J. W. Tukey. The bispectrum and trispectrum were the first generalisations of $f(\omega)$, and these can be considered as respectively a decomposition of skewness and kurtosis over the frequencies, see Brillinger (1984, 1991); Tukey (1959). In general, the basic idea is to Fourier transform the higher order moments or cumulants of (Y_{t+h}, Y_t) instead of the second order moments $\gamma(h)$.

These higher order generalisations of $f(\omega)$ often produce formulas that are hard to estimate, the resulting estimates can be tricky to visualise (e.g. complex-valued), and they can be hard to interpret. The same problems may also occur for the other global and local generalisations of $f(\omega)$, and it is thus advisable to keep in mind the following quote from Akaike (1966) for all the generalisations of $f(\omega)$.

The results of analyses of ordinary spectra and cross-spectra can be understood completely on the basis of linear transformation theory and they suggest the direction of development of models or theories about the phenomena under observation. In contrast to this, higher order spectra seem to be still in want of a sufficiently general theory which gives an overall understandability of them, and their physical meanings have been understood only where a proper model or theory existed before the observation.

It is of interest to note that Brillinger (1965, p. 1372) for higher order spectra gave the following argument in favour of using higher order cumulants instead of higher order moments: ‘The consideration of the cumulant in this [Gaussian] case is not liable to deceive one into believing that he has gained some information. In the non-Gaussian case the cumulant provides

¹This is related to the stochastic coefficients that occur when Y_t is regressed on sines and cosines, i.e. when Y_t is expressed as $Y_t = \int_{-1/2}^{1/2} e^{2\pi i t \omega} dZ(\omega)$, where $Z(\omega)$ is the right continuous orthogonal-increment processes given by the Spectral Representation Theorem, see e.g. Brockwell and Davis (1986, Th. 4.8.2) for details.

an indication of the non-Gaussianity.’ This quote shows that it can be preferable to have a tool that does not trigger any false alarms when the time series under investigation is Gaussian, and this is a key property of the local Gaussian spectral density $f_v(\omega)$.

Alternative (global) generalisations of $f(\omega)$ can be obtained by considering other dependence measures ξ_h based on the random variables Y_t and Y_{t+h} . It is then possible to consider general spectral densities of the form

$$f_\xi(\omega) = \sum_{h \in \mathbb{Z}} \xi_h \cdot e^{-2\pi i \omega h}. \quad (1.3)$$

As noted in (Van Hecke et al., 2017), generalisations of $f(\omega)$ of the form given in eq. (1.3) have been considered in Ahdesmäki et al. (2005) (Kendall’s τ), Carcea and Serfling (2015) (distance correlation) and Zhou (2012) (L -moments).

Hong (1999) introduced a generalised function generalisation of $f(\omega)$, based on the characteristic function. The idea is to transform $\{Y_t\}_{t \in \mathbb{Z}}$ into a u -indexed family of time series based on the characteristic functions, i.e. $\{\exp(iuY_t)\}_{t \in \mathbb{Z}}$, and then consider the bivariate function $\sigma_j(u, v) := \text{Cov}(\exp(iuY_t), \exp(ivY_{t-|j|}))$. This function will be identical to zero for all u and v if and only if Y_t and $Y_{t-|j|}$ are independent, and it is thus possible for this tool to investigate nonlinear time series that are dependent but uncorrelated, like e.g. GARCH-type series. The Fourier transform of $\sigma_j(u, v)$ will exist under some mild regularity assumptions, and it is then possible to define the generalised spectral density function of $\{Y_t\}_{t \in \mathbb{Z}}$ as

$$h(\omega, u, v) := \sum_{h \in \mathbb{Z}} \sigma_j(u, v) \cdot e^{-2\pi i j \omega}, \quad (1.4)$$

The generalised spectral density is in Hong (1999, 2000) used for hypotheses testing and tests of serial dependence (for univariate time series), and recent work has extended the techniques to the case of testing of multivariate time series, see e.g. Li et al. (2016), and a test for conditional independence, see Wang and Hong (2017). This approach has in common with the local Gaussian approach that it is distribution based, not moment based.

Many *local spectral density approaches* have been based on the Fourier transform of *local dependency measures*. Some examples of *local replacements for the autocovariances* $\gamma(h)$ can e.g. be found in Dette et al. (2015), where different *cross-covariance kernels* are defined. In particular, the *Laplace cross-covariance kernel* and *copula cross-covariance kernel* are defined respectively as

$$\gamma_h(x_1, x_2) := \text{Cov}(\mathbb{1}\{Y_{t+h} \leq x_1\}, \mathbb{1}\{Y_t \leq x_2\}), \quad (x_1, x_2) \in \mathbb{R}^2, \quad (1.5a)$$

$$\gamma_h^U(\tau_1, \tau_2) := \text{Cov}(\mathbb{1}\{U_{t+h} \leq \tau_1\}, \mathbb{1}\{U_t \leq \tau_2\}), \quad (\tau_1, \tau_2) \in (0, 1)^2, \quad (1.5b)$$

where $\mathbb{1}\{\cdot\}$ is the indicator function and where knowledge of the marginal distribution G is necessary in order to construct $U_t := G(Y_t)$. Under the assumptions that $\{\gamma_h(x_1, x_2)\}_{h \in \mathbb{Z}}$ and $\{\gamma_h^U(\tau_1, \tau_2)\}_{h \in \mathbb{Z}}$ are absolutely summable, Dette et al. (2015) define the *Laplace* and *copula spectral density kernels* as the corresponding Fourier transformed entities. A rank based Laplace periodogram kernel is also defined.

These spectral density kernels are closely related to the concept of *quantile regression*, introduced in Koenker and Bassett Jr (1978), see also Koenker (2005). Several other alternatives to $f(\omega)$ have been developed based on this concept, like e.g. the *quantilogram* from Linton and Whang (2007), for which the interested reader might consult Han et al. (2016) for more details and additional references. Quantile-based approaches can also be found in Li (2008, 2010a,b,c, 2012a,b, 2014), and in Hagemann (2011). Note that not all of these approaches result in a local tool of the form given in eq. (1.3), see e.g. the *quantile periodogram* from Li (2012c).

The local Gaussian approach to spectral analysis is based on the idea from eq. (1.3), with the variation that ξ_h will be a local measure of (Y_{t+h}, Y_t) that depends on a point $\mathbf{v} = (v_1, v_2)$, where the coordinates v_1 and v_2 corresponds to quantiles of the time series under investigation. The local measure used in this approach is the *local Gaussian autocorrelations* $\rho_{\mathbf{v}}(h)$ from Tjøstheim and Hufthammer (2013), which by construction coincide with the ordinary autocorrelations $\rho(h)$ for Gaussian time series. This implies that the local Gaussian generalisation of $f(\omega)$ by construction will coincide with $f(\omega)$ for Gaussian time series. The *local Gaussian spectral density* $f_{\mathbf{v}}(\omega)$ can thus be used to detect non-Gaussianity in the time series under investigation. This enables an investigation of how different strata of the time series $\{Y_t\}_{t \in \mathbb{Z}}$ interact, and for points on the diagonal, i.e. $v_1 = v_2$, it might then be possible to pick up local periodic phenomena at different scales of the time series, and it could also be used to detect asymmetric behaviour in the lower and upper tails of a time series.

An overview of the paper is as follows: Section 2 defines the *local Gaussian spectral density* $f_{\mathbf{v}}(\omega)$ more precisely and sets up the asymptotic theory for the estimators (the main bulk of the technical details are covered in the appendices). The real and simulated examples in section 3 shows that estimates of $f_{\mathbf{v}}(\omega)$ can be used to detect and investigate nonlinear structures in non-Gaussian white noise, and in particular that $f_{\mathbf{v}}(\omega)$ can detect local periodic phenomena that go undetected in an ordinary spectral analysis. Note that the scripts needed for the reproduction of these examples are contained in the R-package `localgaussSpec`,² where it in addition is possible to use an interactive solution to see how adjustments of the input parameters (used in the estimation algorithms) influence the estimates of $f_{\mathbf{v}}(\omega)$. A discussion is given in section 4, and section 5 presents conclusions.

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 (see also Lacal and Tjøstheim (2017)), and with a comment that these local Gaussian autocorrelations could be used to define a local Gaussian versions of the spectral density from eq. (1.2).

The present section will give a brief summary of the local Gaussian autocorrelations, use these to define the local Gaussian spectral density for strictly³ stationary univariate time series $\{Y_t\}_{t \in \mathbb{Z}}$, and give estimators with a corresponding asymptotic theory.

² Use `devtools::install_github("LAJordanger/localgaussSpec")` to install the package. See the documentation of the function `LG_extract_scripts` for further details.

³ 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 (Y_{t+h}, Y_t) .

2.1 The local Gaussian correlations

The present investigation considers the original concept of the local Gaussian correlation that was given in Tjøstheim and Hufthammer (2013), and it does in addition discuss some modifications of the original definition that will be used later on. Details related to the estimation regime, and asymptotic properties, can be found in appendix B.1.2. 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

The LGC-definition from Tjøstheim and Hufthammer (2013) will now be outlined for the case of a bivariate random variable $\mathbf{W} = (W_1, W_2)$ with joint cdf $G(\mathbf{w})$ and joint pdf $g(\mathbf{w})$. For a specified point $\mathbf{v} := (v_1, v_2)$, the main idea is to find the bivariate Gaussian distribution whose density function best approximates $g(\mathbf{w})$ 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 parameters $\mu_1, \mu_2, \sigma_1, \sigma_2$ and ρ will be denoted by $\boldsymbol{\theta}$,⁴ and the bivariate Gaussian density function will be denoted $\psi(\mathbf{w}; \boldsymbol{\theta})$, i.e.

$$\psi(\mathbf{w}; \boldsymbol{\theta}) := \frac{1}{2\pi \cdot \sigma_1 \sigma_2 \sqrt{1-\rho^2}} \exp \left\{ -\frac{\sigma_1^2(w_1-\mu_1)^2 - 2\sigma_1\sigma_2\rho(w_1-\mu_1)(w_2-\mu_2) + \sigma_2^2(w_2-\mu_2)^2}{2\sigma_1^2\sigma_2^2(1-\rho^2)} \right\}. \quad (2.1)$$

It is natural to require that the following equations are satisfied in order for $\psi(\mathbf{w}; \boldsymbol{\theta})$ to be considered a good approximation of $g(\mathbf{w})$ in a neighbourhood of the point \mathbf{v} ,

$$g(\mathbf{v}) = \psi(\mathbf{w}; \boldsymbol{\theta})|_{\mathbf{w}=\mathbf{v}}, \quad \frac{\partial}{\partial w_1} g(\mathbf{v}) = \frac{\partial}{\partial w_1} \psi(\mathbf{w}; \boldsymbol{\theta}) \Big|_{\mathbf{w}=\mathbf{v}}, \quad \frac{\partial}{\partial w_2} g(\mathbf{v}) = \frac{\partial}{\partial w_2} \psi(\mathbf{w}; \boldsymbol{\theta}) \Big|_{\mathbf{w}=\mathbf{v}}, \quad (2.2)$$

i.e. g and ψ should coincide at \mathbf{v} and they should have coinciding tangent planes there.

It is easy to verify that a solution $\boldsymbol{\theta}$ can be found for any point \mathbf{v} where $g(\mathbf{w})$ is smooth – but these solutions are not unique: $\psi(\mathbf{w}; \boldsymbol{\theta})$ and $\psi(\mathbf{w}; \boldsymbol{\theta}')$ can have coinciding first order linearisation around the point \mathbf{v} , without $\boldsymbol{\theta}$ being identical to $\boldsymbol{\theta}'$. It is possible to extend eq. (2.2) to also include similar requirements for the second order partial derivatives, but the system of equations will then in general have no solution.

To properly account for the higher order terms of $\psi(\mathbf{w}; \boldsymbol{\theta})$ at the point \mathbf{v} , the approximation method needs to include a neighbourhood around \mathbf{v} . Applying the approach used when estimating densities in Hjort and Jones (1996), one can consider a $\mathbf{b} \rightarrow \mathbf{0}^+$ limit of parameters $\boldsymbol{\theta}_b$ that minimise the penalty function

$$q = \int K_b(\mathbf{w} - \mathbf{v}) [\psi(\mathbf{w}; \boldsymbol{\theta}) - g(\mathbf{w}) \log(\psi(\mathbf{w}; \boldsymbol{\theta}))] d\mathbf{w}, \quad (2.3)$$

where $K_b(\mathbf{w} - \mathbf{v})$ is a kernel function with bandwidth \mathbf{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 $g(\mathbf{w})$ and the approximating density $\psi(\mathbf{w}; \boldsymbol{\theta})$. An optimal parameter

⁴The vector $\boldsymbol{\theta}$ is a function of the point \mathbf{v} , but this will be suppressed in the notation.

configuration θ_b for eq. (2.3) should solve the vector equation

$$\int K_b(\mathbf{w} - \mathbf{v}) \mathbf{u}(\mathbf{w}; \boldsymbol{\theta}) [\psi(\mathbf{w}; \boldsymbol{\theta}) - g(\mathbf{w})] d\mathbf{w} = \mathbf{0}, \quad (2.4)$$

where $\mathbf{u}(\mathbf{w}; \boldsymbol{\theta}) := \frac{\partial}{\partial \boldsymbol{\theta}} \log(\psi(\mathbf{w}; \boldsymbol{\theta}))$ is the score function of the approximating density $\psi(\mathbf{w}; \boldsymbol{\theta})$. There will, under suitable assumptions, be a unique limiting solution of eq. (2.4), i.e.

$$\boldsymbol{\theta}_0 = \lim_{b \rightarrow 0^+} \boldsymbol{\theta}_b \quad (2.5)$$

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

Remark 2.1. In the special case where \mathbf{W} is a bivariate normal distributions, i.e. when

$$\mathbf{W} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho \\ \sigma_1 \sigma_2 \rho & \sigma_2^2 \end{bmatrix}\right), \quad (2.6)$$

then, for any point \mathbf{v} and any bandwidth b , the parameters $\boldsymbol{\theta}_b$ that gives the optimal solution of eq. (2.4) will be the parameters given in eq. (2.6). The limit $\boldsymbol{\theta}_0$ in eq. (2.5) will thus of course also be these parameters, which implies that the LGC coincides with the global parameter ρ 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.

2.1.2 Local Gaussian correlation, normalised version

The algorithm that estimates the LGC can run into problems if the data under investigation contains outliers – i.e. the numerical convergence might not succeed for points \mathbf{v} in the periphery of the data. It is possible to counter this problem by removing the most extreme outliers, but that approach might trigger other problems when used on time dependent observations. An alternative strategy based upon normalisation will thus 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 $G(\mathbf{w})$ 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.⁵ Under this continuity assumption, the random variable $\mathbf{W} = (W_1, W_2)$ will have the same copula as the transformed random variable $\mathbf{Z} := (\Phi^{-1}(G_1(W_1)), \Phi^{-1}(G_2(W_2)))$, where Φ is the cdf of the standard normal distribution – whose corresponding pdf as usual will be denoted by ϕ .⁶ This transformed version of \mathbf{W} has standard normal margins, so the LGC-estimation algorithm will not run into numerical problems for this case – which motivates the following alternative approach to the definition of LGC: Instead of finding a Gaussian approximating of the pdf $g(\mathbf{w})$ (of the original random variable \mathbf{W}) at a point \mathbf{v} , find a Gaussian approximation of the pdf

⁵For a proof of this statement, see e.g. Nelsen (2006, Theorem 2.4.3).

⁶See Berentsen et al. (2014b) for an approach where this is used to construct a *canonical local Gaussian correlation* for the copula C .

$g_{\mathbf{z}}(\mathbf{z})$ of the transformed random variable \mathbf{Z} at a transformed point $\mathbf{v}_{\mathbf{z}}$. Expressed relative to the pdf c of the copula C , this means that the strategy in eq. (2.7b) will be used instead of the strategy in eq. (2.7a).

$$g(\mathbf{w}) = c(G_1(w_1), G_2(w_2)) g_1(w_1) g_2(w_2) \quad \text{approximate at } \mathbf{v} = (v_1, v_2), \quad (2.7a)$$

$$g_{\mathbf{z}}(\mathbf{z}) = c(\Phi(z_1), \Phi(z_2)) \phi(z_1) \phi(z_2) \quad \text{approximate at } \mathbf{v}_{\mathbf{z}} := (\Phi^{-1}(G_1(v_1)), \Phi^{-1}(G_2(v_2))). \quad (2.7b)$$

The normalised 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 \mathbf{W} is bivariate Gaussian. The transformed random variable \mathbf{Z} corresponding to the \mathbf{W} from eq. (2.6) will then be $\mathbf{Z} = ((W_1 - \mu_1)/\sigma_1, (W_2 - \mu_2)/\sigma_2)$, which implies

$$\mathbf{Z} \sim \mathbf{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right), \quad (2.8)$$

so the normalised LGC will thus also coincide with the global parameter ρ at all points.

2.1.3 Local Gaussian correlation, simplified normalised version

The numerical estimation of the normalised LGC, based on eq. (2.7b), avoids by construction the numerical convergence problems that can occur for the estimates of the general version in eq. (2.7a). The analysis of the convergence rate of the normalised LGC must take into account that there is an additional *normalise the margins* step, but this does not affect the convergence rate, see remark 2.7, page 10 for further details and references.

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_b in eq. (2.3). Briefly summarised, the 5×5 covariance matrix of the estimate $\hat{\theta}_b$ will have the form $V_b^{-1} W_b V_b^{-1}$, the presence of the kernel K_b means that the matrices V_b and W_b have rank one in the limit $b \rightarrow \mathbf{0}^+$, and this slows down the convergence rate, cf. Tjøstheim and Hufthammer (2013, Th. 3) for the details.

The property that the limiting matrices have rank one does not pose a problem when only one parameter is estimated,⁷ and the convergence rate would then be much faster (i.e. $\sqrt{nb_1 b_2}$). Inspired by the fact that the transformed random variable \mathbf{Z} have standard normal margins, it has been introduced a simplified normalised version of the LGC where only the ρ -parameter should be estimated when using the approximation approach from eq. (2.7b), i.e. the values of μ_1, μ_2 are taken to be 0, whereas σ_1^2 and σ_2^2 are taken to be 1. This simplified approach has been applied successfully with regard to density estimation⁸ in Otneim and Tjøstheim (2016, 2017), and it thus seems natural to also include this approach in this paper too.

The algorithm used to find the optimal value for the single parameter ρ follows the same recipe as the one used when five parameters are present, so the theoretical framework is unchanged. Moreover, from the discussion around eq. (2.8), it is clear that the simplified LGC version also gives the correct answer when the random variable \mathbf{W} itself is Gaussian. However, the simplified LGC will in general deviate from the normalised LGC-version in section 2.1.2 –

⁷The matrices then becomes 1×1 , so the singularity problems does not occur.

⁸Note that it is *not* the local Gaussian correlation that is the target of interest when this simplified approach is used for density estimation, as will be discussed in more detail in appendix C.6.

and in fact, it might be regions where none of the desired properties listed in eq. (2.2) holds when the simplified version is used. The geometric intuition from the general case can thus not be applied when working with the simplified approach, cf. appendix C.6 for a more detailed discussion.

2.2 The local Gaussian spectral densities

An extension of eq. (1.2) can in principle be based on any of the three LGC that was encountered in sections 2.1.1 to 2.1.3, but (in order to avoid the aforementioned numerical convergence problems) only the latter two of them will be considered here, i.e. the time series will be normalised before local Gaussian autocorrelations are computed.

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

- (a) *With G the univariate marginal cumulative distribution of $\{Y_t\}_{t \in \mathbb{Z}}$, and Φ the cumulative distribution of the standard normal distribution, define a normalised version $\{Z_t\}_{t \in \mathbb{Z}}$ of $\{Y_t\}_{t \in \mathbb{Z}}$ by*

$$\{Z_t := \Phi^{-1}(G(Y_t))\}_{t \in \mathbb{Z}}. \quad (2.9)$$

- (b) *For a given point $\mathbf{v} = (v_1, v_2)$ and for each lag $h \neq 0$ bivariate pair $\mathbf{Z}_{h:t} := (Z_{t+h}, Z_t)$, a local Gaussian autocorrelation $\rho_{\mathbf{v}|p}(h)$ can be computed, where the p specifies if the correlations stems from a one or five parameter approximation of the bivariate marginal density of $\mathbf{Z}_{h:t}$ at (v_1, v_2) . The convention $\rho_{\mathbf{v}|p}(0) \equiv 1$ is used when $h = 0$, since no bivariate density is present for this case.*

- (c) *When $\sum_{h \in \mathbb{Z}} |\rho_{\mathbf{v}|p}(h)| < \infty$, the local Gaussian spectral density at the point \mathbf{v} is defined as*

$$f_{\mathbf{v}|p}(\omega) := \sum_{h=-\infty}^{\infty} \rho_{\mathbf{v}|p}(h) \cdot e^{-2\pi i \omega h}. \quad (2.10)$$

Remark 2.2. The requirement $\sum_{h \in \mathbb{Z}} |\rho_{\mathbf{v}|p}(h)| < \infty$ in definition 2.1(c) 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 $\mathbf{v} \in \mathbb{R}^2$.

The following definition of time reversible time series, from Tong (1990, def. 4.6), is needed in lemma 2.3(c).

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

Lemma 2.3. *The following properties holds for $f_{\mathbf{v}|p}(\omega)$.*

- (a) *$f_{\mathbf{v}|p}(\omega)$ coincides with $f(\omega)$ for all $\mathbf{v} \in \mathbb{R}^2$ when $\{Y_t\}_{t \in \mathbb{Z}}$ is a Gaussian time series, and when $\{Y_t\}_{t \in \mathbb{Z}}$ consists of i.i.d. observations.*

(b) The following holds when $\check{\mathbf{v}} := (v_2, v_1)$ is the diagonal reflection of $\mathbf{v} = (v_1, v_2)$;

$$f_{\mathbf{v}|p}(\omega) = 1 + \sum_{h=1}^{\infty} \rho_{\check{\mathbf{v}}|p}(h) \cdot e^{+2\pi i \omega h} + \sum_{h=1}^{\infty} \rho_{\mathbf{v}|p}(h) \cdot e^{-2\pi i \omega h}, \quad (2.11a)$$

$$f_{\mathbf{v}|p}(\omega) = \overline{f_{\check{\mathbf{v}}|p}(\omega)}. \quad (2.11b)$$

(c) When $\{Y_t\}_{t \in \mathbb{Z}}$ is time reversible, then $f_{\mathbf{v}|p}(\omega)$ is real valued for all $\mathbf{v} \in \mathbb{R}^2$, i.e.

$$f_{\mathbf{v}|p}(\omega) = 1 + 2 \cdot \sum_{h=1}^{\infty} \rho_{\mathbf{v}|p}(h) \cdot \cos(2\pi \omega h). \quad (2.12)$$

(d) $f_{\mathbf{v}|p}(\omega)$ will in general be complex-valued, but it will always be real valued when the point \mathbf{v} lies on the diagonal, i.e. when $v_1 = v_2$. Equation (2.12) will hold in this diagonal case too.

Proof. Item (a) follows for the Gaussian case since the local Gaussian autocorrelations $\rho_{\mathbf{v}|p}(h)$ 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 becomes 1. Items (b) to (d) are trivial consequences of the diagonal folding property from lemma C.1, i.e. $\rho_{\mathbf{v}|p}(-h) = \rho_{\check{\mathbf{v}}|p}(h)$, and the definition of time reversibility, see appendices C.1 and C.2 for details. \square

Remark 2.3. For general points $\mathbf{v} = (v_1, v_2)$, the complex valued result of $f_{\mathbf{v}|p}(\omega)$ might be hard to investigate and interpret – but, due to lemma 2.3(d), the investigation becomes simpler for points on the diagonal.⁹ The real valued results $f_{\mathbf{v}|p}(\omega)$ for \mathbf{v} along the diagonal can be compared with the result of the ordinary (global) spectral density $f(\omega)$, as given in eq. (1.2), 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 $f_{\mathbf{v}|p}(\omega)$ might be interpreted as indicators of e.g. *periodicities at a local level*. This implies that estimates of $f_{\mathbf{v}|p}(\omega)$ might be useful as an exploratory tool, an idea that will be pursued in section 3.

Remark 2.4. Note that the collection of local Gaussian autocorrelations $\{\rho_{\mathbf{v}|p}(h)\}_{h \in \mathbb{Z}}$ might not be non-negative definite. Caution is thus advised if peaks and troughs of $f_{\mathbf{v}|p}(\omega)$ are attempted interpreted as they would have been if they had occurred for an ordinary (global) spectral density $f(\omega)$. See the discussion in section 4 for further details.

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. This FFT-approach can not be used in the local case since there is no natural factorisation 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

⁹A diagonal point corresponds to a situation where observations of the same ‘scale of magnitude’ are compared.

This can in particular be of interest for time series featuring an asymmetric behaviour, since a comparison of the local Gaussian spectra at points corresponding to e.g. the 10% and 90% quantiles might (as seen in fig. 11) reveal nonlinear structures which the ordinary spectral density $f(\omega)$ fails to detect.

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.

Definition 2.4. For a sample $\{y_t\}_{t=1}^n$ of size n , an m -truncated estimate $\widehat{f}_{\mathbf{v}|p}^m(\omega)$ of $f_{\mathbf{v}|p}(\omega)$ are constructed by means of the following procedure.

- (a) Find an estimate \widehat{G}_n of the marginal cumulative distribution function, and compute the pseudo-normalised observations $\{\widehat{z}_t := \Phi^{-1}(\widehat{G}_n(y_t))\}_{t=1}^n$ that corresponds to $\{y_t\}_{t=1}^n$.
- (b) Create the lag h pseudo-normalised pairs $\{\widehat{z}_{t+h}, \widehat{z}_t\}_{t=1}^{n-h}$ for $h = 1, \dots, m$, and estimate, both for the point $\mathbf{v} = (v_1, v_2)$ and its diagonal reflection $\check{\mathbf{v}} = (v_2, v_1)$, the local Gaussian autocorrelations $\{\widehat{\rho}_{\mathbf{v}|p}(h|b_h)\}_{h=1}^m$ and $\{\widehat{\rho}_{\check{\mathbf{v}}|p}(h|b_h)\}_{h=1}^m$, where the $\{b_h\}_{h=1}^m$ is the bandwidths used during the estimation of the local Gaussian autocorrelation for the different lags.
- (c) Adjust eq. (2.11a) from lemma 2.3(b) with some lag-window function $\lambda_m(h)$ to get the estimate

$$\widehat{f}_{\mathbf{v}|p}^m(\omega) := 1 + \sum_{h=1}^m \lambda_m(h) \cdot \widehat{\rho}_{\mathbf{v}|p}(h|b_h) \cdot e^{+2\pi i \omega h} + \sum_{h=1}^m \lambda_m(h) \cdot \widehat{\rho}_{\check{\mathbf{v}}|p}(h|b_h) \cdot e^{-2\pi i \omega h}. \quad (2.13)$$

The selection of bandwidth and truncation level is discussed in sections 4.1 and 4.2.

The following result is an analogue to eq. (2.12) of lemma 2.3(c)

Lemma 2.5. 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 $\widehat{f}_{\mathbf{v}|p}^m(\omega)$ can for all points $\mathbf{v} \in \mathbb{R}^2$ be written as

$$\widehat{f}_{\mathbf{v}|p}^m(\omega) = 1 + 2 \cdot \sum_{h=1}^m \lambda_m(h) \cdot \widehat{\rho}_{\mathbf{v}|p}(h|b_h) \cdot \cos(2\pi \omega h). \quad (2.14)$$

Moreover, eq. (2.14) will always hold when the point \mathbf{v} lies on the diagonal, i.e. $v_1 = v_2$.

Proof. This follows from items (c) and (d) of lemma 2.3. □

Remark 2.5. The estimated \widehat{G}_n in definition 2.4(b) can e.g. be the (rescaled) empirical cumulative distribution function created from the sample $\{y_t\}_{t=1}^n$, or it could be based on some log spline technique like the one implemented in Otneim and Tjøstheim (2016).

Remark 2.6. The bandwidths $\mathbf{b}_h = (b_{h1}, b_{h2})$ in definition 2.4(b) does not need to be equal for all the lags h when an estimate $\widehat{f}_{\mathbf{v}|p}^m(\omega)$ is computed. For the asymptotic investigation it is sufficient to require that b_{h1} and b_{h2} approach zero at the same rate, i.e. that there exists $\mathbf{b} = (b_1, b_2)$ such that $b_{hi} \asymp b_i$ for $i = 1, 2$ and for all h (that is to say, $\lim b_{hi}/b_i = 1$).

Remark 2.7. The asymptotic theory for $\widehat{\rho}_{\mathbf{v}|p}(h|b_h)$, given that the required regularity conditions are satisfied, follows from Otneim and Tjøstheim (2016); Tjøstheim and Hufthammer (2013). The analysis in (Tjøstheim and Hufthammer, 2013) considered the general case with a $p = 5$ parameter local Gaussian approximation at the point \mathbf{v} for the lag h pairs $\{(y_{t+h}, y_t)\}_{t=1}^{n-h}$, i.e. the original observations $\{y_t\}_{t=1}^n$ were used instead of the normalised observations $\{z_t := \Phi^{-1}(G(y_t))\}_{t=1}^n$. Since the cumulative density function G in general will be unknown, the present asymptotic

analysis must work with the pseudo-normalised observations $\{\widehat{z}_t\}_{t=1}^n$, which makes it necessary to take into account the difference between the true normalised values z_t and the estimated pseudo-normalised values \widehat{z}_t . The analysis in (Otneim and Tjøstheim, 2016) revealed that $\widehat{G}_n(y_t)$ approaches $G(y_t)$ at a faster rate than the rate of convergence for the estimated local Gaussian correlation, so the convergence rate of $\widehat{\rho}_{v|p}(h|b_h)$ will thus not be affected by the distinction between z_t and \widehat{z}_t . The present analysis will not duplicate the arguments related to this distinction, and the interested reader should consult (Otneim and Tjøstheim, 2016, Section 3) for the details.

Remark 2.8. The bias-variance balance of the m -truncated estimates $\widehat{f}^m(\omega)$ of the ordinary spectral density $f(\omega)$ depends on the size of m relative to n (the size of the sample). The bias-variance balance for the estimates $\widehat{f}_{v|p}^m(\omega)$ must in addition consider the size of m relative to both n and the bandwidths $\{b_h\}_{h=1}^m$, i.e. 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 v , i.e. whether the point v lies at the center or in the periphery of the pseudo-normalised observations $\{(\widehat{z}_{t+h}, \widehat{z}_t)\}_{t=1}^{n-h}$. Confer section 3.2 for further details.

Figure 1 shows the effect of the pseudo-normalisation on the dmbp example¹⁰ 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-normalised transformation of it, and it is clear that the shape of the pseudo-normalised version resembles the shape of the original version. The effect of the transformation is twofold; it removes the extreme outliers at the same time as it spreads out the center of the distribution.

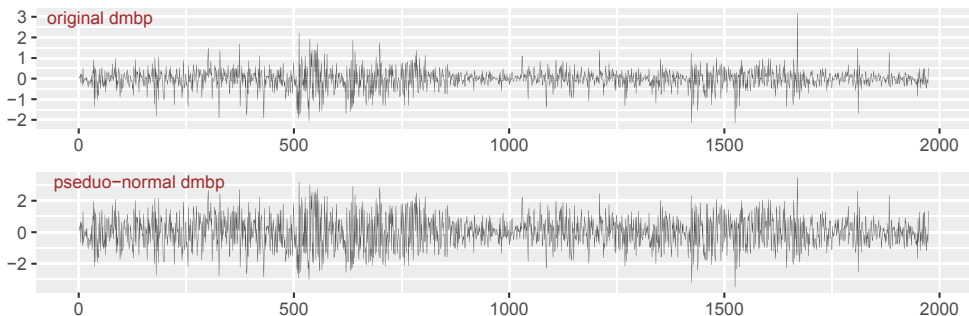


Figure 1: dmbp, original version and pseudo-normalised version.

2.4 Asymptotic theory for $\widehat{f}_{v|p}^m(\omega)$

This section presents asymptotic results for the cases where $\widehat{f}_{v|p}^m(\omega)$ are real-valued functions. Note that both assumptions and results are stated relative to the original observations instead of

¹⁰ 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 (2015b), where the following description was given: ‘The daily percentage nominal returns computed as $100 [\ln(P_t) - \ln(P_{t-1})]$, where P_t is the bilateral Deutschemark/British pound rate constructed from the corresponding U.S. dollar rates.’

the pseudo-normalised observations. This simplification does not affect the final convergence rates (see remark 2.7 for details) and it makes the analysis easier. The requirement that the LGSD should be defined relative to the normalised 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 $\{Y_t\}_{t \in \mathbb{Z}}$ requires components related to the bivariate lag h pairs that can be constructed from it, whereas the theoretical analysis of $\widehat{f}_{v|p}^m(\omega)$ also requires that $(m+1)$ -variate pairs are considered. Note that item (c) of definition 2.4 implies that it is sufficient to only consider positive values for h .

Definition 2.6. 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,

$$\mathbf{Y}_{h:t} := [Y_{t+h}, Y_t]', \quad \mathbf{Y}_{\bar{m}:t} := [Y_{t+m}, \dots, Y_t]', \quad (2.15)$$

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

Remark 2.9. The densities g_h are those needed when investigating the local Gaussian estimates for the different lags h . The bivariate densities g_h can all be obtained from the $(m+1)$ variate density $g_{\bar{m}}$ by integrating out the $m-1$ redundant marginals, which in particular implies that if an $(m+1)$ -variate function $\tilde{\eta}_h(\mathbf{y}_{\bar{m}}) : \mathbb{R}^{m+1} \rightarrow \mathbb{R}^1$ is the *obvious extension*¹¹ of a bivariate function $\eta_h(\mathbf{y}_h) : \mathbb{R}^2 \rightarrow \mathbb{R}^1$, then

$$\mathbb{E}[\eta_h(\mathbf{Y}_{h:t})] = \mathbb{E}[\tilde{\eta}_h(\mathbf{Y}_{\bar{m}:t})], \quad \text{for } h \in \{1, \dots, m\}. \quad (2.16)$$

With the notation from definition 2.6 the following assumption can now be imposed upon Y_t . Note that items (e) to (g) contains references to definitions that first are given in appendix B; these definitions are related to the estimation of a penalty function – 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 random variables X_{hqi}^n in item (g) in essence are the result that occurs when the product of the kernel function $K_{h:b}(\mathbf{y}_h - \mathbf{v})$ and the score function of the local Gaussian approximation $\psi(\mathbf{y}_h)$ is evaluated in $\mathbf{y}_h = \mathbf{Y}_{h:t}$.

Assumption 2.1. The univariate process $\{Y_t\}_{t \in \mathbb{Z}}$ will be assumed to satisfy the following properties, with $\mathbf{v} = (v_1, v_2)$ in item (d) the point at which the estimate $\widehat{f}_{v|p}^m(\omega)$ of $f_{v|p}(\omega)$ are to be computed.

- (a) $\{Y_t\}_{t \in \mathbb{Z}}$ is strictly stationary.

¹¹The obvious extension is to consider the function to be a constant with respect to all the new variables that are introduced.

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

$$\sum_{j=1}^{\infty} j^a [\alpha(j)]^{1-2/\nu} < \infty \quad \text{for some } \nu > 2 \text{ and } a > 1 - 2/\nu. \quad (2.17)$$

(c) $\text{Var}(Y_t^2) < \infty$.

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

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

$$g_h(\mathbf{y}_h) = g_h(\mathbf{v}) + \mathbf{g}_h(\mathbf{v})' [\mathbf{y}_h - \mathbf{v}] + \mathfrak{R}_h(\mathbf{y}_h)' [\mathbf{y}_h - \mathbf{v}], \quad (2.18)$$

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

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

(e) There exists a bandwidth \mathbf{b}_{h0} such that there for every $\mathbf{0} < \mathbf{b} < \mathbf{b}_{h0}$ is a unique minimiser $\boldsymbol{\theta}_{h:b}$ of the penalty function $q_{h:b}$ defined in eq. (B.4), page 43.

(f) The collection of bandwidths $\{\mathbf{b}_{h0}\}_{h \in \mathbb{Z}}$ has a positive infimum, i.e. there exists a \mathbf{b}_0 such that

$$\mathbf{0} < \mathbf{b}_0 := \inf_{h \in \mathbb{Z}} \mathbf{b}_{h0}, \quad (2.19)$$

which implies that this \mathbf{b}_0 can be used simultaneously for all the lags.

(g) For $X_{hq:i}^n$ from definition B.11, page 51, the bivariate, trivariate and tetravariate density functions must be such that the expectations $\mathbb{E}[X_{hq:i}^n]$, $\mathbb{E}[|X_{hq:i}^n|^\nu]$ and $\mathbb{E}[X_{hq:i}^n \cdot X_{jr:k}^n]$ all are finite.

Remark 2.10. These assumption upon Y_t are extensions of those used for the LGC-case in Tjøstheim and Hufthammer (2013). Assumption 2.1(b) 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 extends to the present case.

Remark 2.11. The α -mixing requirement in item (b) ensures that Y_{t+h} and Y_t will be asymptotically independent as $h \rightarrow \infty$, i.e. the bivariate density functions $g_h(\mathbf{y}_h)$ will for large lags h approach the product of the marginal densities, and the situation will thus stabilise when h is large enough. This is in particular of importance for item (f), since it implies that it will be possible to find a nonzero \mathbf{b}_0 that works for all h .

Remark 2.12. The finiteness requirements in assumption 2.1(g) will be trivially satisfied if the densities are bounded, i.e. they will then be consequences of properties of the kernel function K_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 $u(w; \theta)$ of $\psi(w; \theta)$

The score function in eq. (2.4), i.e. $u(w; \theta) := \frac{\partial}{\partial \theta} \log(\psi(w; \theta))$, plays a central role in the local density-estimation approach of Hjort and Jones (1996), and it thus 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 $\widehat{\theta}_v - \theta_v$ does implicitly require that $u(v; \theta_v) \neq \mathbf{0}$ in order for the corresponding asymptotic covariance matrix to be well defined. The investigation of $(\widehat{f}_{v|p}^m(\omega) - f_{v|p}(\omega))$ in this paper builds upon the asymptotic results from Tjøstheim and Hufthammer (2013), and the following assumption must thus be satisfied in order for the given convergence rates and asymptotic variances to be valid. Note that the index p as usual does show whether it is a one or a five parametric local Gaussian approximation $\psi_p(w; \theta_p)$ that is considered, and that $u_p(w; \theta_p)$ here represent the corresponding score function.

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

- (a) $u_p(v; \theta_{p|v}(h)) \neq \mathbf{0}$ for all finite h .
- (b) $\lim u_p(v; \theta_{p|v}(h)) \neq \mathbf{0}$.

Remark 2.13. It is, for a given time series Y_t and a given point v , possible to inspect the p equations in $u_p(w; \theta_p) = \mathbf{0}$ in order to see when items (a) and (b) of assumption 2.2 might fail to hold true. It is e.g. possible to find the parameter-configurations $\theta_{p|v}^\dagger$ which solve $u_p(v; \theta_p) = 0$, and then observe that assumption 2.2(a) will fail if $\theta_{p|v}^\dagger \in \{\theta_{p|v}(h)\}$. For the case of the asymptotic requirement in item (b), the key observation is that the strong mixing requirement from assumption 2.1(b) implies that Y_{t+h} and Y_t will become independent when $h \rightarrow \infty$. Together with the assumption of normalised marginals, this implies that the limit of $\theta_{p|v}(h)$ always becomes $[\mu_1, \mu_2, \sigma_1, \sigma_2, \rho]^\dagger = [0, 0, 1, 1, 0]^\dagger$, which means that assumption 2.2(b) will fail for any point v that solves $u_p(v; [0, 0, 1, 1, 0]^\dagger) = 0$.

Remark 2.14. The one parameter local Gaussian approximation $\psi_1(w; \theta_1)$ is less flexible than the five parameter approximation $\psi_5(w; \theta_5)$, and this lack of flexibility can for some time series Y_t imply that assumption 2.2(a) is bound to fail at some points v , see the discussion in appendix C.6.2 for further details.

2.4.3 Assumptions for n , m and b

For simplicity, the present analysis will use the $\mathbf{b} = (b_1, b_2)$ introduced in remark 2.6, see page 10, i.e. it will be assumed that the individual bandwidths 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. For the present case, where the lag h pairs are of the form (Y_{t+h}, Y_t) it might also be natural to assume that b_1 and b_2 should approach zero at the same rate, i.e. that $b_1 \asymp b_2$, but this will not be imposed from the outset.

Assumption 2.3. Let $m := m_n \rightarrow \infty$ be a sequence of integers denoting the number of lags to include, and let $\mathbf{b} := \mathbf{b}_n \rightarrow \mathbf{0}^+$ be the bandwidths used when estimating the local Gaussian correlations for the lags $h = 1, \dots, m$ (based on n observations). Let b_1 and b_2 refer to the two components of \mathbf{b} , and let α, ν and a be as introduced in assumption 2.1(b). Let $s := s_n \rightarrow \infty$ be a sequence of integers such that $s = o\left(\sqrt{nb_1 b_2 / m}\right)$, and let τ be a positive constant. The

following requirements must be satisfied for these entities.¹²

- (a) $\log n/n(b_1 b_2)^5 \rightarrow 0$, (only required for the case $p = 5$).
- (b) $nb_1 b_2/m \rightarrow \infty$.
- (c) $m^\delta (b_1 \vee b_2) \rightarrow 0$, where $\delta = 2 \vee \frac{\nu(a+1)}{\nu(a-1)-2}$.
- (d) $\sqrt{nm/b_1 b_2} \cdot s^\tau \cdot \alpha(s - m + 1) \rightarrow \infty$.
- (e) $m = o((nb_1 b_2)^{\tau/(2+5\tau)-\lambda})$, for some $\lambda \in (0, \tau/(2 + 5\tau))$.
- (f) $m = o(s)$.

Remark 2.15. Assumption 2.3(a) is needed for the case $p = 5$ in order for the asymptotic theory from Tjøstheim and Hufthammer (2013) to be valid for the estimates $\widehat{\rho}_{v|5}(h)$.

Remark 2.16. See lemma C.3 for a verification of the internal consistency of the requirements given in assumption 2.3.

2.5 Convergence theorems for $\widehat{f}_{v|p}^m(\omega)$

Theorem 2.7 (v on diagonal, i.e. $v_1 = v_2$). *The local Gaussian spectral density $f_{v|p}(\omega)$ is a real valued function when the point v lies on the diagonal. Furthermore; when the univariate time series Y_t satisfies assumptions 2.1 and 2.2, and n , m and $\mathbf{b} = (b_1, b_2)$ are as given in assumption 2.3, then the following asymptotic results holds for the m -truncated estimate $\widehat{f}_{v|p}^m(\omega)$,*

$$\sqrt{n(b_1 b_2)^{(p+1)/2}/m} \cdot \left(\widehat{f}_{v|p}^m(\omega) - f_{v|p}(\omega) \right) \xrightarrow{d} N(\mathbf{0}, \sigma_{v|p}^2(\omega)), \quad (2.20)$$

where the formula

$$\sigma_{v|p}^2(\omega) = 4 \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{h=1}^m \lambda_m^2(h) \cdot \cos^2(2\pi\omega h) \cdot \widetilde{\sigma}_{v|p}^2(h) \quad (2.21)$$

relates the variance $\sigma_{v|p}^2(\omega)$ to the asymptotic variances $\widetilde{\sigma}_{v|p}^2(h)$ of $\sqrt{n(b_1 b_2)^{(p+1)/2}} \cdot (\widehat{\rho}_{v|p}(h|\mathbf{b}_n) - \rho_{v|p}(h))$.

Proof. The proof is given in appendix A.1. □

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

Theorem 2.8 (Y_t time reversible). *The local Gaussian spectral density $f_{v|p}(\omega)$ is a real valued function for all points v when Y_t is time reversible (see definition 2.2, page 8). Furthermore under assumptions 2.1 to 2.3, the same asymptotic results as stated in theorem 2.7 holds for the m -truncated estimate $\widehat{f}_{v|p}^m(\omega)$.*

Proof. Lemma 2.3(c) states that $f_{v|p}(\omega)$ is a real-valued function, and the proof of theorem 2.7 (see appendix A.1) can then be repeated without any modifications. □

¹²Notational convention: ‘ \vee ’ denotes the maximum of two numbers, whereas ‘ \wedge ’ Denotes the minimum.

Remark 2.17. The asymptotic normality results in theorems 2.7 and 2.8 does not easily enable a computation of confidence intervals for the estimated LGSD. Thus, the 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 has 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.

Remark 2.18. The asymptotic result for $\widehat{f}_{v|p}^m(\omega)$ complex-valued is given in appendix A.2, where it can be seen that $\sqrt{n(b_1 b_2)^{(p+1)/2}/m} \cdot \left(\widehat{f}_{v|p}^m(\omega) - f_{v|p}(\omega) \right)$ then asymptotically approaches a complex-valued normal distribution instead of a real-valued one.

3 Examples and possible interpretations

This section will investigate if the m -truncated estimates $\widehat{f}_{v|p}^m(\omega)$ might have a potential as an exploratory tool. It will be verified that it does behave as expected for the cases where it is known what the result should be (i.e. Gaussian time series), that it can detect the presence of local structures (including periodicities) in a constructed example, and finally, that it enables a visual aid to see how good a GARCH-type model fitted to the dmbp-data¹³ seems to match the result from the data themselves. Note that the discussion of some of the technical details are postponed to section 4.

Remark 3.1. All the simulated time series investigated in this section have the same length as the dmbp-series, i.e. they all have length 1974. This common length seems like a natural restriction to apply for this first investigation of $\widehat{f}_{v|p}^m(\omega)$ as an exploratory tool, since different lengths otherwise could be an explanation for any observed differences. The estimation machinery does produce similar results for shorter time series too, but it remains to be investigated how long a time series ought to be in order to avoid that small sample variation distorts the signal of any local structures that might be present.

Remark 3.2. The same reasoning as in remark 3.1 motivates that the configuration of the input parameters will be kept the same for the different cases to be investigated, see section 3.1 for details about the selected values.

Remark 3.3. The initial step of the computation of $\widehat{f}_{v|p}^m(\omega)$ is to replace the observations $\{y_t\}_{t=1}^n$ with the corresponding pseudo-normal observations $\{\widehat{z}_t\}_{t=1}^n$, cf. definition 2.4, i.e. an estimate of the marginal cumulative density function G is needed. The present analysis has used the rescaled empirical cumulative density function \widehat{G}_n for this purpose, but the computations could also have been based on a logspline-estimate of G , see remark 2.5. For the time series investigated in this section, a preliminary investigation indicated that the two normalisation procedures created strikingly similar estimates of $\widehat{f}_{v|p}^m(\omega)$, so the computationally faster approach based on the rescaled empirical cumulative density-function has been applied for the present investigation.

Remark 3.4. The estimation of $\widehat{f}_{v|p}^m(\omega)$ does also include the selection of p , i.e. whether a 5-parameter or a 1-parameter local Gaussian approximation should be used. As noted in appendix C.6, the 1-parameter local Gaussian approximation might be useful when estimating

¹³See footnote 10 (page 11) for a description of the dmbp-data.

densities, but the estimated parameters $\widehat{\rho}_{v|1}(h)$ might not give a good indicator of the local dependency-structure of the targeted distribution. The 5-parameter estimates $\widehat{\rho}_{v|5}(h)$ have thus been selected for all of the plots, with the sole exception of fig. 14 where two plots based on $\widehat{\rho}_{v|1}(h)$ have been included to emphasise why it is best to avoid them in this context.

Remark 3.5. The *pointwise confidence bands*¹⁴ shown in the plots later on are all based upon $R = 100$ replicates. Repeated independent samples from the known model was used to construct the confidence bands in section 3.3, whereas block-bootstrap was used for the real data example in section 3.4. The lower and upper limits of the pointwise confidence bands are based on the 0.05 and 0.95 quantiles of the resulting collection of estimated local Gaussian spectral densities (truncated at lag m), and thus gives an estimated 90% pointwise confidence band for $f_{v|p}^m(\omega)$.

3.1 Setting the input parameters

Several input parameters must be selected before an estimate of $\widehat{f}_{v|p}^m(\omega)$ can be obtained. The main parameters that must be taken into account are listed below, with the values that will be used for the examples later on. A further discussion of some of these items are given in section 4.

1. v , the points to investigate, will for the present investigation be diagonal points whose coordinates corresponds to the 10%, 50% and 90% percentiles of the standard normal distribution, i.e. the values are -1.28, 0 and 1.28. Information about the point of investigation is contained in the upper right corner of the relevant plots, where it is marked as 10% : 10%, and so on. The corresponding coordinates are $(-1.28, -1.28)$, $(0, 0)$ and $(1.28, 1.28)$, and these will often be referred to as *lower tail*, *center* and *upper tail* when discussed in the text.
2. ω , the frequencies to investigate. Values between 0 and $\frac{1}{2}$.
3. $\mathbf{b} = (b_1, b_2)$, the bandwidth-vector to be used when computing the local Gaussian autocorrelations. Most of the plots shown in this section have used $\mathbf{b} = (.5, .5)$, with the exception of fig. 6, where plots based on $\mathbf{b} = (.75, .75)$ and $\mathbf{b} = (1, 1)$ have been included for comparison.¹⁵
4. m , the truncation level, i.e. the number of lags to include in the estimate of $\widehat{f}_{v|p}^m(\omega)$. The value $m = 10$ has been used in this investigation, and this number is by default given in the upper left corner of the relevant plots.
5. $\lambda_m(h)$, the weighting function to be used for the smoothing of the different lags. The Tukey-Hanning lag-window kernel has been used for all the present examples, i.e.

$$\lambda_m(h) = \begin{cases} \frac{1}{2} \cdot (1 + \cos(\pi \cdot \frac{h}{m})) & |h| \leq m, \\ 0 & |h| > m. \end{cases}$$

Remark 3.6. The R-package `localgaussSpec` can be used for the estimation of $\widehat{f}_{v|p}^m(\omega)$ for a wide combination of alternatives for the parameters,¹⁶ and it allows an integrated interactive

¹⁴The pointwise confidence band gives for each frequency ω a confidence interval for the value of $f_{v|p}^m(\omega)$.

¹⁵It is natural to require $b_1 = b_2$ since both of the components in the lag h pseudo-normalised pairs comes from the same univariate time series.

¹⁶See footnote 2 (page 4) for details about installation of the `localgaussSpec`-package.

investigation of the results by means of a shiny-application.¹⁷ Note that the R-package contains all the scripts needed for the exact recreation of the plots included in this section.

Remark 3.7. The R-package `localgauss`, see Berentsen et al. (2014a), was used for the estimation of the local Gaussian auto-correlations for the $p = 5$ case. These estimates are returned with an indicator (named `eflag`) that reveals whether or not the estimation algorithm converged numerically to the estimate, and this numerical convergence-information has then been added to the relevant plots in their lower left corner. In particular, ‘NC = OK’ will be used to show that all the required estimates had a successful numerical convergence. Contrary, ‘NC = FAIL’ will represent that problems did occur during the estimation algorithm. It should be noted that convergence-problems hardly occurs when the computations are based on pseudo-normalised observations.

Remark 3.8. It has to be admitted upfront that there is an unresolved issue with regard to the selection of the blocklength for the bootstrapping of the `dmbp`-example shown in fig. 11, see the discussion in section 4 for further details.

3.2 Estimation aspects for the given parameter configuration

The estimation of $\hat{f}_{v|5}^m(\omega)$ for a point $\mathbf{v} = (v_1, v_2)$ that lies on the diagonal, i.e. $v_1 = v_2$, will be based on the estimates of $\hat{\rho}_{v|5}(h)$ for $h = 1, \dots, m$. It is thus of interest to investigate how the estimates $\hat{\rho}_{v|5}(h)$ depends on the parameter-configuration given in section 3.1.

First of all, note that the combination of point \mathbf{v} and bandwidth \mathbf{b} influences how many of the h -lagged pairs that effectively contributes to the computation of $\hat{\rho}_{v|5}(h)$. This is shown in fig. 2 where the pseudo-normalised `dmbp`-data (of length 1974) has been used as an example. In the plot of the pseudo-normalised time series, the three horizontal dashed lines represent the *levels* which corresponds to the coordinates of the three points \mathbf{v} , whereas the horizontal strips centered at those lines shows which observations that lies within a distance of $b = 0.5$ from the respective lines. The three plots at the bottom shows the corresponding 1-lagged pairs, each with a *bandwidth-square* (of width $2b$) centered at one of the selected points \mathbf{v} .

Remark 3.9. In fig. 2, the *bandwidth-strip* at the center of the trajectory plot contains 756 observations, whereas the two other strips both contains 355 observations. Note that the bandwidth-strips for the tails must contain the same number of observations due to the symmetry enforced by the pseudo-normalisation, and furthermore note that all time series of this particular length will end up with pseudo-normalised trajectories that (for the given combination of points \mathbf{v} and bandwidth \mathbf{b}) must have the exact same number of observations inside of their bandwidth-strips as those encountered here.

Remark 3.10. In order for an h -lagged pseudo-normalised pair $(\hat{z}_{t+h}, \hat{z}_t)$ to occur within a lag h bandwidth-square (centered at a diagonal point \mathbf{v}), it is necessary that both \hat{z}_{t+h} and \hat{z}_t lie inside the corresponding bandwidth-strip. For the case $h = 1$, shown at the bottom of fig. 2, the number of points inside the three bandwidth-squares thus counts how many neighbouring pseudo-normalised observations that occurred in the respective bandwidth-strips. The number of observations captured in the three $h = 1$ cases are respectively 75, 359 and 66, and several comments can be based on these numbers. First of all, these numbers indicates that there might

¹⁷ See Chang et al. (2017) for details about shiny.

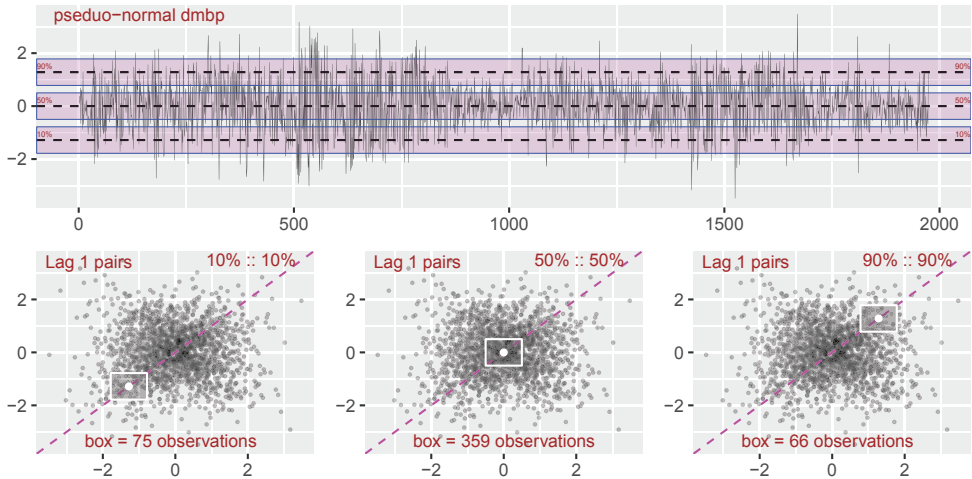


Figure 2: dmbp (pseudo-normalised version), *levels* and *bandwidth-bands* (top) and *lag 1 bandwidth-squares* (bottom).

be an asymmetry between the lower and upper tails of the dmbp-data. Furthermore, as the bias-variance properties of the estimates $\hat{\rho}_{v|5}(h)$ depends on the number of points that effectively contributes during the computation, it is clear that the variance will increase for points 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.

Remark 3.11. An important detail with regard to the estimation of $f_{v|5}(\omega)$ is the selection of the truncation level m , since that value (in addition to the value of the bandwidth \mathbf{b}) influences the bias-variance properties of the estimate $\hat{f}_{v|5}^m(\omega)$. It would be preferable if some data driven method could be used to identify an optimal range of values within which m should lie, or at least have some rule of thumb that could be used during an investigation. An initial approach might be to apply some existing rule of thumb used for the selection of m for the m -truncated estimates of the ordinary spectral density $f(\omega)$, but it remains to be investigated whether or not that would give a reasonable truncation level when estimating $f_{v|5}(\omega)$.

Remark 3.12. If the truncation level m is too large, the interconnection between m and \mathbf{b} could create a situation (for points v in the periphery of the data) where the number of lag h pseudo-normalised observations used to estimate $\hat{\rho}_{v|5}(h)$ might become too small to give a reasonable estimate. It seems likely that it will be a difficult task to construct a general selection method for the truncation level m , but it is not hard to investigate (before any estimates are produced) how many pseudo-normalised lag h pairs that for a given combination of h , v and \mathbf{b} lies inside of the corresponding bandwidth-square. For the dmbp-example it can e.g. be noted that the number of lag h pseudo-normalised pairs that occurs inside a given bandwidth-square will fluctuate a bit as h increases, but that it obviously must decrease as h grows larger (since the total number of lag h pairs decreases linearly). The numbers of such pseudo-normalised pairs that occurs within the bandwidth-squares for the $h = 200$ version of fig. 2 are respectively 70, 263 and 63, which for this particular case does not seem to represent a drastic decrease in the number of pseudo-normalised observations that are available in the tails. This does of course not imply

that an estimate of $f_{v|5}(\omega)$ based on a truncation level of $m = 200$ will necessarily make sense in the dmbp-case (for the present points of interest), but it could be used as an indicator that a higher truncation level could have been applied than the one used later on.

Figure 3 shows how $\widehat{\rho}_{v|5}(h)$ varies for the three points of interest (when $\mathbf{b} = (0.5, 0.5)$). Red dotted lines shows the truncation level $m = 10$ (to be used later on), in order to emphasise which estimates of $\widehat{\rho}_{v|5}(h)$ that will contribute to the estimation of $\widehat{f}_{v|5}^m(\omega)$. This plot shows that there is a clear distinction between the center and the two tails. The $\widehat{\rho}_{v|5}(h)$ tends to fluctuate around 0 at the center, which implies that the corresponding estimated spectral density $\widehat{f}_{v|5}^m(\omega)$ most likely will be rather flat and close to 1. For the two tails, it seems natural to assume that some long-range dependency must be present, and one might also suspect that there is an asymmetry between the two tails.

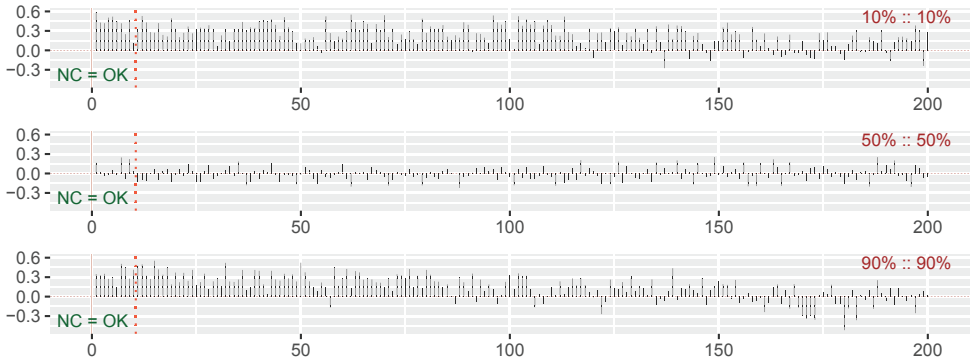


Figure 3: dmbp-data, $\widehat{\rho}_{v|5}(h)$ for $h = 1, \dots, 200$ (for the three points of interest).

The cumulative sums of the autocorrelations from fig. 3, are presented in fig. 4, and once more the plot indicates an asymmetry between the two tails. Moreover, since a long initial sequence of positive values (a bit larger than zero) for $\widehat{\rho}_{v|5}(h)$ automatically implies that the m -truncated estimated spectral density $\widehat{f}_{v|5}^m(\omega)$ must have a peak for the frequency $\omega = 0$, it follows from fig. 4 that the local Gaussian spectral densities at the tail-points must have such peaks at $\omega = 0$. These details are easier to see in fig. 11, page 29, where $\widehat{f}_{v|5}^m(\omega)$ for the dmbp-data are presented.

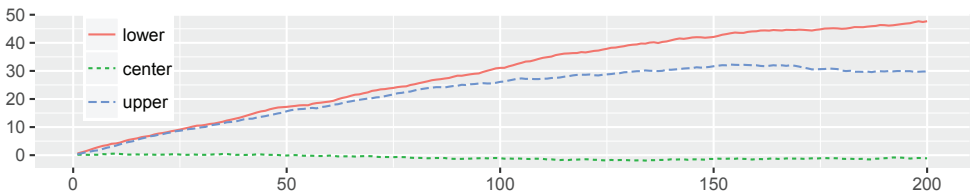


Figure 4: dmbp-data, cumulative sum of local Gaussian auto-correlations.

3.3 Some simulations

This section will estimate the local Gaussian spectral densities $f_{v|5}(\omega)$ for simulated data. It is known from lemma 2.3(a) that $f_{v|5}(\omega)$ coincides with the ordinary (global) spectral density

$f(\omega)$ when the time series under investigation either is i.i.d. or Gaussian. This can be used to test the sanity of the estimation algorithm, since repeated estimates based on independent realisations from these models should be distributed around the expected value if the algorithm works as intended.

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 $\widehat{f}_{v|p}^m(\omega)$ and $\widehat{f}^m(\omega)$ for each of the replicates. Then extract the median of these estimates to get estimates of the (m -truncations of the) true values, and select suitable upper and lower percentiles of the estimates to produce an estimate of the pointwise confidence bands. Finally, create plots that contains estimates and pointwise confidence bands for the m -truncated versions of $f_{v|p}(\omega)$ and $f(\omega)$, see the definition below.

Definition 3.1. The m -truncated versions $f_{v|p}^m(\omega)$ and $f^m(\omega)$ of $f_{v|p}(\omega)$ and $f(\omega)$, for a specified weighting function $\lambda_m(h)$, is defined by means of

$$f_{v|p}^m(\omega) := 1 + \sum_{h=1}^m \lambda_m(h) \cdot \rho_{v|p}(h) \cdot e^{+2\pi i \omega h} + \sum_{h=1}^m \lambda_m(h) \cdot \rho_{v|p}(h) \cdot e^{-2\pi i \omega h}, \quad (3.1a)$$

$$f^m(\omega) := \sum_{h=-m}^m \lambda_m(h) \cdot \rho(h) \cdot e^{-2\pi i \omega h}. \quad (3.1b)$$

3.3.1 Gaussian white noise

Figure 5 shows the result when the estimation procedure is used on 100 independent samples of length 1974 from a standard normal distribution $N(0, 1)$. The computations are based on the bandwidth $\mathbf{b} = (0.5, 0.5)$, 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-normalised version of the first time series that was sampled from the model, with dashed brown lines at the levels that corresponds to the above mentioned points. The three other panels contains information about the m -truncated ordinary spectral density $f^m(\omega)$ (red part, the same for all the plots) and the m -truncated local Gaussian spectral densities $f_{v|p}^m(\omega)$ for the three points under investigation (blue part). Information about the truncation level and the points are printed at the top of each plot.

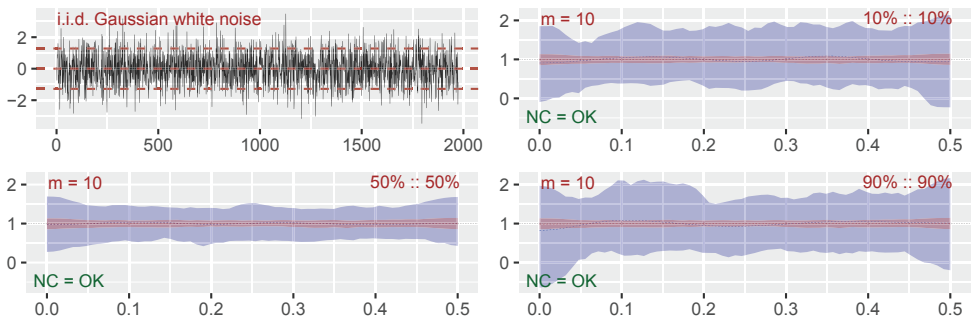


Figure 5: i.i.d. Gaussian white noise.

It can be seen from fig. 5 that the medians of the estimates (the red and blue dashed lines) are good estimates of $f^m(\omega)$ and $f_{v|5}^m(\omega)$ (the m -truncations of the true values), which in this case in fact coincides with $f(\omega)$ and $f_{v|5}(\omega)$, i.e. 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 bands 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 confidence bands are wider in the tails, which is a natural consequence of the reduced number of points in those region, see the discussion related to fig. 2.

Remark 3.13. The estimation procedure gave good estimates of the true values $f(\omega)$ and $f_{v|5}(\omega)$ in the simple example of fig. 5, but it is important to keep in mind that these plots actually shows estimates of $f^m(\omega)$ and $f_{v|5}^m(\omega)$. It might be necessary to apply a (much) higher truncation level before $f^m(\omega)$ and $f_{v|5}^m(\omega)$ gives decent approximations of the true values $f(\omega)$ and $f_{v|5}(\omega)$. It thus seems preferable to estimate $\hat{f}_{v|5}^m(\omega)$ 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.

3.3.2 Some trigonometric examples

The Gaussian white noise example in fig. 5 shows that the estimated local Gaussian spectral density behaved in the anticipated manner for that simple case, but it is of interest to see if the result looks reasonable for other examples too. However, beyond the realm of Gaussian time series, it is not known what the true value for the local Gaussian spectral density actually should be – which poses a problem for such an investigation. This section will thus construct a *local trigonometric time series* for which it at least can be reasonably argued what the expected outcome should be for some specially designated points v (given a suitable bandwidth b). These artificial time series will in general not satisfy the requirements needed for the asymptotic theory (both in the global and local case) to hold true, but they can still be used to investigate if an exploratory tool based on the local Gaussian spectral density might reveal 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 necessary to first investigate the result based on independent samples from a time series of the form $Y_t = \cos(2\pi\alpha t + \phi) + w_t$, where w_t is Gaussian white noise with mean zero and standard deviation σ , and where it in addition is such that α is fixed for all the replicates whereas the phase-adjustment ϕ is randomly generated for each individual replicate. A realisation with $\alpha = 0.302$ and $\sigma = 0.75$ is shown in fig. 6, where the frequency α has been indicated with a vertical line in order to show that both the local and global approach 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 fig. 6, before the *local trigonometric* case is defined and investigated.

Remark 3.14. All the plots in fig. 6 shows the same point 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 $b = (.5, .5)$ (the bandwidth used in all the other examples), whereas the two plots at the bottom shows the situation for the bandwidths $b = (.75, .75)$ and $b = (1, 1)$, respectively at

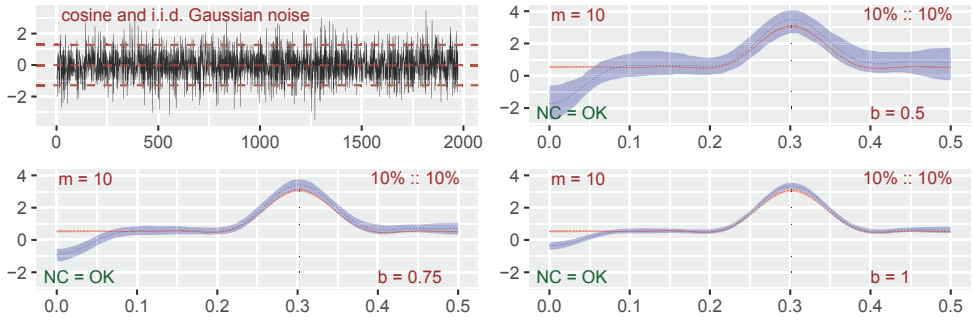


Figure 6: Single cosine and i.i.d. white noise, same point, bandwidths based on 0.5, 0.75 and 1.

the left and right. In this case, the widths of the pointwise confidence-bands are influenced by the selected bandwidths, but the overall shape seems to be similar. This feature is also present for the other examples that have been investigated.

Remark 3.15. Based on the width of the pointwise confidence bands in fig. 6, one might wonder if the bandwidth $\mathbf{b} = (.5, .5)$ is too small, and that it perhaps would be better to use a larger bandwidth. However, it is important to keep in mind that the bandwidth is used in a kernel function $K_{\mathbf{b}}$ that estimates $\rho_{v|5}(h)$, and when the bandwidth grows this estimate will converge to the estimate of the ordinary global autocorrelation – which would make it impossible to detect any local effects.

Remark 3.16. For the example in fig. 6, it will be a large difference in the plots when more lags are included, i.e. the peaks will grow taller and narrower. However, the position of the peaks will not move, and that indicates that these plots (even for rather low truncation values) might reveal some properties of the underlying structure. Again, this feature is shared with the other examples that have been investigated.

Remark 3.17. The local Gaussian spectral densities in fig. 6 goes below zero for low frequencies, a feature that is not entirely unexpected as $\{\rho_{v|5}(h)\}_{h \in \mathbb{Z}}$, the collection of local Gaussian autocorrelations, may not be a non-negative definite function. In fact, based on the observation that the estimates of $\widehat{f}_{v|5}^m(\omega)$ have peaks that are taller and wider than those of $\widehat{f}^m(\omega)$, 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 $\widehat{f}_{v|5}^m(\omega)$ thus requires that it has to lie below the estimates of $\widehat{f}^m(\omega)$ in some other region, and if necessarily it must attain negative values somewhere. The interesting details in the plots are thus the position of the peaks of $\widehat{f}_{v|5}^m(\omega)$, and regions with negative values should not in general be considered a too troublesome feature.

The local trigonometric case: The next case to be investigated is an artificially constructed model where different *local* cosines are next used to create a process close to white noise, see the top-panel of fig. 7 for a realisation. The basic recipe for these time series use the following simple principle: For a given $r \geq 2$, select a collection of different base levels (L_1, \dots, L_r) at the y -axis, a collection of amplitudes (A_1, \dots, A_r) , a collection of frequencies $(\alpha_1, \dots, \alpha_r)$ and a collection of phase-adjustments (ϕ_1, \dots, ϕ_r) . Finally, assign a probability p_i to each

$i = 1, \dots, r$, such that $\sum_{i=1}^r p_i = 1$. In order to allow more randomness into the sample, it is also possible to specify an additional amplitude adjustment (A'_1, \dots, A'_r) . The amplitude will, for each t , be selected uniformly from the interval spanned by A_i and A'_i when both are specified, and this uniformly random amplitude function will then be denoted $A_i(t)$. (Note that $A_i(t) \equiv A_i$ if A'_i is unspecified.)

The preceding ingredients enables the definition of the following functions,

$$C_i(t) = L_i + A_i(t) \cdot \cos(2\pi\alpha_i t + \phi_i), \quad i = 1, \dots, r, \quad (3.2)$$

from which a stochastic variable Y_t can be created by means of the probabilities (p_1, \dots, p_r) , i.e. let N_t be a random variable that with probability p_i takes the value i , and define

$$Y_t := \sum_{i=1}^r C_i(t) \cdot \mathbb{1}\{N_t = i\}, \quad (3.3)$$

where the indicator function $\mathbb{1}\{\}$ ensures that only one of the $C_i(t)$ contribute at a given value t . Note that it is assumed that the phases ϕ_i are uniformly drawn (one time for each realisation) from the interval between 0 and 2π , and that it moreover also is assumed that the stochastic processes ϕ_i , $A_i(t)$ and N_t are independent of each other. Based on this, the autocovariance of Y_{t+h} and Y_t can be given as a function of L_i and p_i , from which it then is fairly easy to select a combination of input parameters that returns a Y_t -process that looks like white noise.

The time series presented here has $r = 4$ components with base levels L_i in $(-2, -1, 0, 1)$, amplitude-functions $A_i(t)$ defined by A_i in $(1.0, 0.5, 0.3, 0.5)$ and A'_i in $(0.5, 0.2, 0.2, 0.6)$, and frequencies α_i in $(0.267, 0.091, 0.431, 0.270)$. For this case the probabilities p_i in $(0.05, 0.28, 0.33, 0.33)$ was used to sample¹⁸ which component to include in Y_t .

Figure 7 shows a *simplified* excerpt of length 100 from one realisation of Y_t , where $A_i(t) \equiv A_i$ in order to emphasise which one of underlying ‘hidden’ components $C_i(t)$ (shown as dotted curves) that was selected in this case (the phase-adjustments ϕ_i in this particular realisation are $(0.52, 2.57, 3.24, 2.49)$). Note that the amplitudes A_i for this example was selected to give a minimal level of overlap between the ranges of the functions $C_i(t)$. The center panel of fig. 7 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 behaves almost like white noise. Note that the vertical lines in the center panel shows the frequencies α_i that was used in eq. (3.2).

The bottom panel of fig. 7 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 \mathbf{v} (given a suitable bandwidth \mathbf{b}). First of all, the bottom panel shows *one* of the cosines from the top panel, the red circles represents the points from the top panel that happened to lie on this particular cosine – and the blue crosses represents all the remaining points (at integer valued times t) of the cosine. Recall that these points are from the simplified realisation where $A_i(t) \equiv A_i$, and that the actual values thus would be distorted a bit due to additional randomisation from the amplitude adjustments A'_i .

¹⁸The printed probabilities might not add to one! This is due to the fact that these values was rounded in R before they were included in this document by the means of the R-package `knitr`, see (Xie, 2015, 2016) for details about dynamic documents.

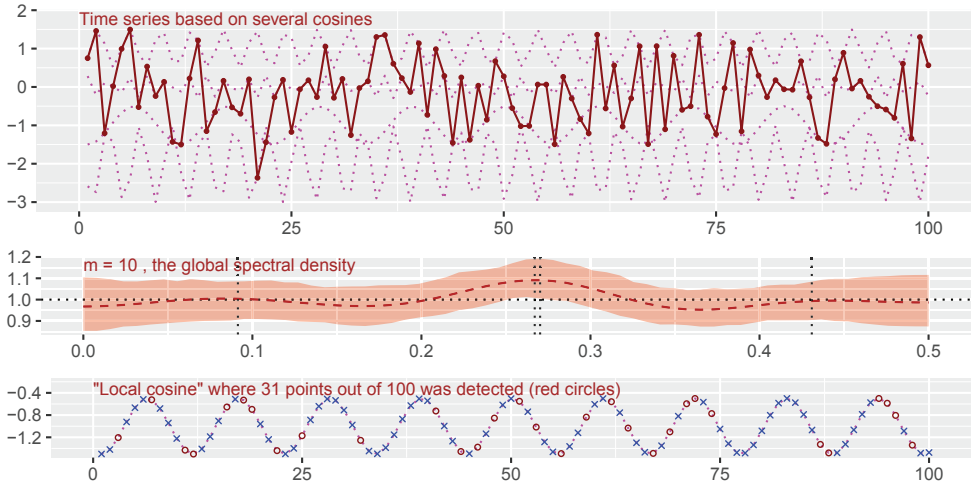


Figure 7: Top: Short excerpt from artificial example based on *hidden trigonometric components*. Center: Estimated (truncated) global spectral density (*hidden frequencies* indicated with vertical lines). Bottom: Local cosine showing the detected points at the local level centered at -1 .

The red circles can be considered as a randomly selected collection of points from a time series like the one investigated in fig. 6, and the main point of interest is that it (for a sufficiently long time series, and a sufficiently large bandwidth \mathbf{b}) will be the case that the estimated local Gaussian auto-correlations 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 \mathbf{v}) is computed from the sets of bivariate points (Y_{t+h}, Y_t) . In particular: It might not have a detrimental effect upon 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 red points should be fairly close to the situation shown in fig. 6, at least if the time series under investigation is sufficiently long.

Given this heuristic argument, and the observation that the input parameters used in eq. (3.3) gives time series whose pseudo-normalised traces will have their 10%-, 50%- and 90%-quantiles approximately corresponding to the original levels L_i in $(-1, 0, 1)$, it can be postulated that the estimated local Gaussian spectral densities at the designated points 10% : 10%, 50% : 50% and 90% : 90%, all should resemble fig. 6 – with peaks at the respective frequencies α_i in $(0.091, 0.431, 0.270)$.

The local investigation by means of $\hat{f}_{\mathbf{v}|15}^m(\omega)$ is presented in fig. 8, as usual based on 100 independent samples of length 1974, a bandwidth $\mathbf{b} = (0.5, 0.5)$ and showing 90% pointwise confidence bands. This shows that an exploratory tool based on the (m -truncated) local Gaussian spectral density indeed is capable of detecting the expected peaks at the designated points mentioned in the preceding discussion. In particular, $\hat{f}_{\mathbf{v}|15}^m(\omega)$ picks up different peaks at different points, and these peaks changes quite a bit from the lower tail to the upper tail.

Remark 3.18. It should be noted that this simple example was created with a combination of L_i , A_i and p_i that gave peaks approximately at the three points investigated in this section, and that the plots for other points might vary quite a bit. It is thus of importance to investigate a range of points and check if/how the shape of $\widehat{f}_{v_{15}}^m(\omega)$ changes as the point varies from the lower tail to the upper tail. Note also that the rather low value for p_1 implies that the $C_1(t)$ seems to go undetected. An investigation of the local behaviour for this component would require a point at a lower quantile than the present value for the lower tail, and it seems likely that an investigation at such a point might run into problems due to a scarcity of observations in the vicinity of the point.

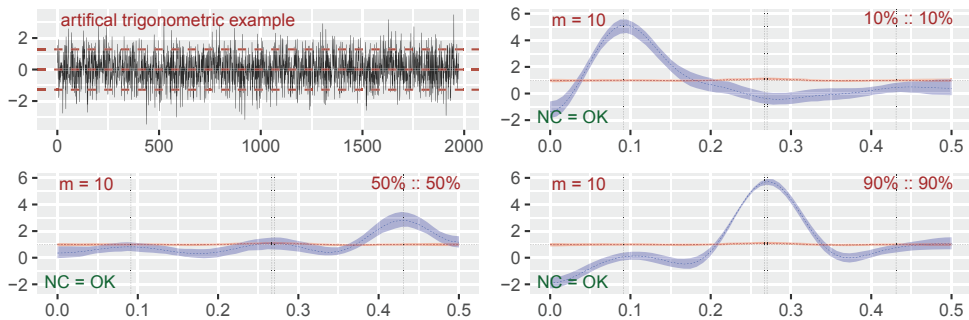


Figure 8: Artificial example, *hidden trigonometric components*.

3.3.3 Beware of global structures

It is important to keep in mind that a comparison of the local Gaussian spectral density $f_{v_{15}}(\omega)$ and the ordinary spectral density $f(\omega)$ can reveal deviations from the property of the time series Y_t being *Gaussian* – and for time series whose ordinary (global) spectrum looks like white noise, this can be interpreted as a *detection of nonlinear traits* in the time series under investigation.

For time series with a non-flat global spectrum, it might not be a good idea to automatically consider a difference between $f_{v_{15}}(\omega)$ and $f(\omega)$ as a sign of nonlinear traits, as shown in fig. 9 where a more *extreme version* of the case investigated in fig. 6 are presented. In this case 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 $Y_t = \cos(2\pi\alpha t + \phi) + w_t$, where $\alpha = 0.302$ (as before), whereas the standard deviation of the Gaussian white noise w_t has been reduced to $\sigma = 0.05$.

The low value of the standard deviation σ implies that samples from this time model have a very clear periodic behaviour, as can be seen from the plots in fig. 9, 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. 9, where the 100 first pseudo-normalised observations of one of the samples are presented.

The main detail of interest in fig. 9 is the clear deviation between the local and global spectra, as seen for the points 10% : : 10% and 50% : : 50% at the truncation level $m = 10$ and for 10% : : 10% at $m = 20$. Figure 9 reveals the importance of including both the local and global spectra in the investigation, and it shows that caution should be exercised when trying to in-

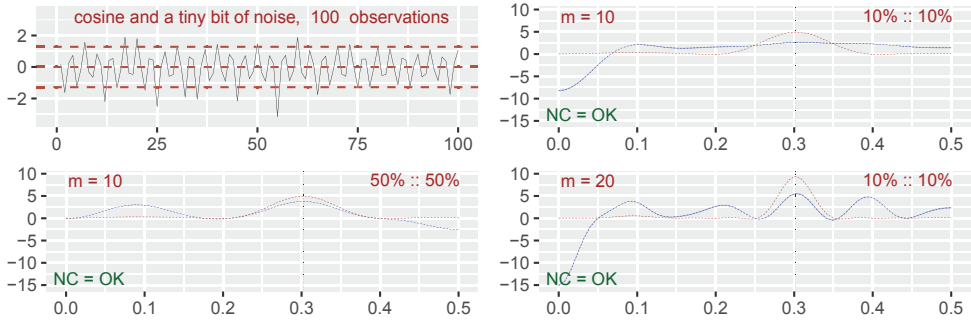


Figure 9: Pseudo-normalised single cosine and a tiny bit of noise.

terpret a difference between local and global spectra for a time series with a non-flat global spectral density.

It might be natural (for such cases) to proceed with an approach where some model is fitted to the data, preferably one that gives residuals that (globally) behaves like white noise, and then perform a new local Gaussian analysis upon the residuals. For the present case of investigation, that could in principle detect local information hidden in the white noise that was added on top of the trigonometric function.

3.3.4 A GARCH-type model

The next example is a GARCH-type model, more precisely an *asymmetric power ARCH-model* (apARCH) of order $(2, 3)$, with parameters based on a fitting to the dmbp-data. (The R-package rugarch, Ghalanos (2015b) was used to find the parameters of several GARCH-models, and the asymmetric power ARCH model with the best fit was then selected.) The apARCH (p, q) model was introduced in Ding et al. (1993), where it was given as

$$\epsilon_t = s_t e_t, \quad e_t \sim \mathbf{N}(0, 1), \quad (3.4a)$$

$$s_t^\delta = \alpha_0 + \sum_{i=1}^p \alpha_i (|\epsilon_{t-i}| - \gamma_i \epsilon_{t-i})^\delta + \sum_{j=1}^q \beta_j s_{t-i}^\delta, \quad \text{where} \quad (3.4b)$$

$$\alpha_0 > 0, \delta \geq 0, \quad \alpha_i \geq 0, i = 1 \dots, p, \quad -1 < \gamma_i < 1, i = 1 \dots, p, \quad \beta_j \geq 0, j = 1 \dots, q, \quad (3.4c)$$

but the apARCH $(2, 3)$ -model used in this example is a bit more complicated than the one from (Ding et al., 1993), see Ghalanos (2015a, sec. 2.2.5), for the details.

Figure 10 shows the result from a local Gaussian investigation of the above mentioned apARCH-model, as usual showing 90% pointwise confidence bands constructed from 100 independent samples of length 1974, and with a bandwidth $\mathbf{b} = (0.5, 0.5)$. The m -truncated ordinary global spectral density $f^m(\omega)$ of a GARCH-type model like the one investigated here is known to be 1 (since $\rho(h) = 0$ when $h \neq 0$), and fig. 10 shows that the estimate of $f^m(\omega)$ indeed is close to 1. These plots do in addition indicate that the estimated $f_{v_{15}}^m(\omega)$ differs a lot from $f^m(\omega)$ in the tails, but not in the center. The question now is whether or not the shape of these (m -truncated) local Gaussian spectral densities might reveal anything about the behaviour of the time series at the levels corresponding to the points.

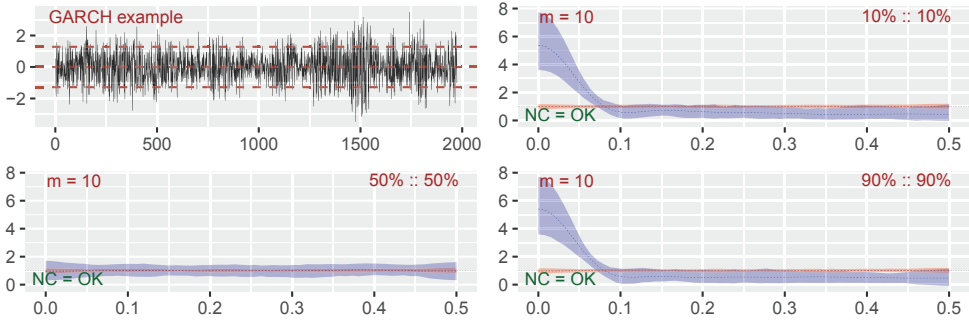


Figure 10: GARCH-type model, based on dmbp.

For the lower and upper tails, the present example seems to indicate a symmetric situation, and it seems to be ample reason to claim that the local values $f_{v|5}^m(\omega)$ for low frequencies are significantly different from the global values $f^m(\omega)$ when the pointwise confidence bands are taken into account.

It is clear from lemma 2.5 that $\hat{f}_{v|5}^m(0) = 1 + 2 \cdot \sum_{h=1}^m \lambda_m(h) \cdot \hat{\rho}_{v|5}(h)$, so the peaks observed at the lower and upper tails thus reveals that the first batch of estimated local Gaussian autocorrelations consists of a sequence of positive values – which indicates that long range dependencies might be present. This impression is strengthened when plots with a higher truncation level is considered, as the peak at the frequency $\omega = 0$ continues to grow. Compare fig. 3 to see the situation for the estimated local Gaussian autocorrelations $\hat{\rho}_{v|5}(h)$ for the dmbp-data, upon which the parameters of the apARCH-model was based. Furthermore, the fact that $f_{v|5}^m(\omega)$ seems to be very close to 1 at the center indicates that the estimated local Gaussian autocorrelations $\hat{\rho}_{v|5}(h)$ at the center fluctuates around 0, which again is in agreement with the impression fig. 3 gives with regard to the dmbp-data at this point.

3.4 Real data

The data to be used in the present section will be the Bollerslev-Ghysel benchmark data set (dmbp), see footnote 10 (page 11) for details. A plot of the pseudo-normalised dmbp-data was given in fig. 2, and the estimates $\hat{\rho}_{v|5}(h)$ was investigated in fig. 3. The apARCH(2,3)-model used to create fig. 10 had parameters obtained from a fitting to the dmbp-data, and the present investigation will reveal that estimates of the m -truncated local Gaussian spectral density can provide some visual aid with regard to the quality of the tested GARCH-type model – in particular, this might be of interest when doing model selection.

The estimation of $\hat{f}_{v|5}^m(\omega)$ for a given point v at a given frequency ω requires a selection of a bandwidth b and some maximum number of the lags m , and these will be kept the same as those used for the simulated data, since that seems to be the natural option when a comparison of the corresponding estimated local Gaussian spectral densities is of interest. The number of replicates used to create the confidence bands will likewise be kept the same, but issues related to the resampling strategy for the given sample (see discussion below) might have an effect upon that part.

Figure 11 presents the results based upon the dmbp-data. Note that this plot differs a bit from those encountered for the simulated data; a solid line represents the estimate from the ac-

tual (length 1974) sample at hand, using the bandwidth $\mathbf{b} = (0.5, 0.5)$, and the 90% pointwise confidence band was constructed from estimates based on 100 resampled versions of the original data. The resampling was done by means of a block-bootstrap, where the selection of the blocklength (in this case 100) turned out to be a problem – since it (to the best of the authors’ knowledge) does not exist a method that can return a data-driven value for the blocklength to be used for a sample from a nonlinear time series with a flat spectrum – see the discussion in section 4 for further details.

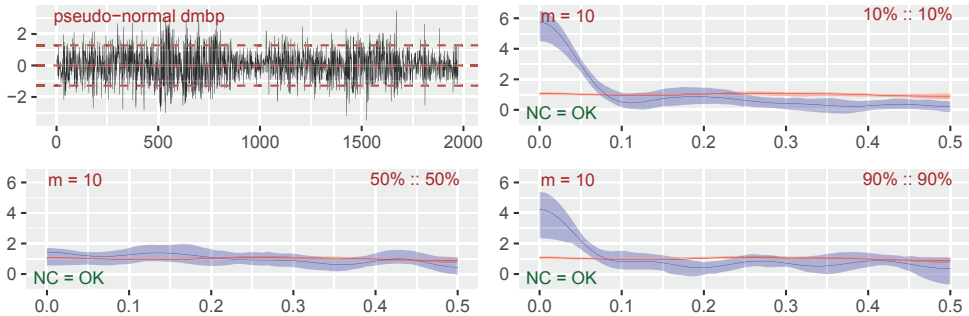


Figure 11: dmbp-data, bootstrapped based confidence intervals.

The solid line in the two right panels in fig. 11 indicates an asymmetry between the lower and upper tails, which seems natural when one takes into account the observations from figs. 3 and 4 that the local Gaussian autocorrelations for lagged pairs at the lower tail tends to have higher values than those occurring at the upper tail – and this is in agreement with the asymmetry between a *bear market* (going down) and a *bull market* (going up). In particular, note that the detected long range behaviour (peak at $\omega = 0$) for the extremes are more prominent for the for bear market than for the bull market.

Keeping in mind that the apARCH(2, 3)-model from fig. 10 had coefficients obtained from a fit to the dmbp-data, it is of interest to compare the dashed lines from that plot with the solid lines in the present plot. Since the asymmetry observed in fig. 11 are missing from fig. 10, it could be that the present GARCH-type model might not be optimal for the dmbp. But it has to be emphasised that the pointwise confidence intervals in figs. 10 and 11 are rather wide, so it might be premature to reject the apARCH(2, 3)-model as an adequate model for the dmbp data. *Remark 3.19.* It was a problem to figure out which blocklength to use in the block-bootstrap algorithm. The plots in fig. 11 used the value 100 for the blocklength, and this was selected after some tests with different blocklengths. The reason for the selection of this blocklength was partially due to the impression from fig. 3 that a long block length might be needed, and it was partially due to a desire for the original estimate (i.e. the solid line) to be approximately at the center of the estimated confidence band – and it has to be noted that such an approach could lead to erroneous conclusions. However, it should also be noted that the significant difference (in the tails, for frequencies ω between 0 and 0.07) between local and global m -truncated spectral densities still was present when other blocklengths was tested, so it seems safe to conclude from fig. 11 that local properties have been detected in the dmbp-data for the lower and upper tails. Note that the lower-tail part of fig. 3 in fact could indicate that an even larger blocklength than 100 should be used, see section 4 for further details.

3.5 Inspecting the local Gaussian autocorrelations

It might be enlightening to not only compare the estimated local Gaussian spectral densities, but also to compare the estimated local Gaussian auto-correlations. Figure 12 illustrates this by showing the first 20 lags for three different dmbp-related cases. The top panel shows the estimates based on the pseudo-normalised dmbp-data, the center panel shows box-plots based on the 100 bootstrapped replicates (using the block-length 100), whereas the bottom panel shows box-plots based on the 100 samples (of length 1974) from the apARCH(2,3)-model that was fitted to the original dmbp-data. The point under investigation is 10% : 10%, and the estimates are (as before) based on the bandwidth $\mathbf{b} = (0.5, 0.5)$.

A comparison of such plots of local Gaussian autocorrelations might be useful with regard to the problem of judging the appropriateness of a proposed block-length for the bootstrapping procedure, and it might also be possible to detect if a model fitted to the data clearly fails to mimic the local behaviour of the data the model was fitted to.

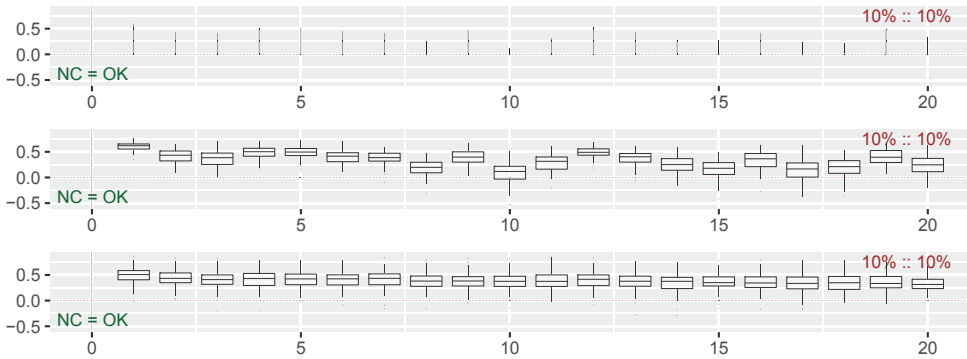


Figure 12: Local Gaussian auto-correlations related to the pseudo-normalised dmbp-data. Top: Original data. Center: Bootstrapped results (100 replicated using the blocklength 100). Bottom: apARCH(2,3) fitted to dmbp (100 simulations).

3.6 Exploration for off-diagonal points

The preceding examples all considered points $\mathbf{v} = (v_1, v_2)$ on the diagonal, i.e. $v_1 = v_2$, but mathematically it is as such not a problem to estimate $f_{v|v}^m(\omega)$ for points off the diagonal, see appendix A.2 for the relevant asymptotic theory. However, the estimates becomes complex-valued when $v_1 \neq v_2$, which makes them harder to visualise and interpret – and there is no clear level interpretation as in fig. 2. Nevertheless, when $\hat{f}_{v|v}^m(\omega)$ is used as an exploratory tool in relation to model selection, it could still be of interest to compare plots based on a model fitted to the data and the plots based on the original data. As a follow up of the preceding investigation, this section will compare the apARCH(2,3)-model and the dmbp-data at the point $\mathbf{v} = (-1.28, 1.28)$, i.e. the first coordinate corresponds to the 10% quantile and the second coordinate corresponds to the 90% quantile.

Since the estimates $\hat{f}_{v|v}^m(\omega)$ now are complex-valued, it seems natural to investigate them by using plots based on their corresponding real and imaginary parts (cartesian presentation) or plots based on their amplitude and phase (polar representation). This section will use the

cartesian presentation, where both the real and complex parts are investigated at the same time, as shown in fig. 13.

In the cartesian case, the estimate can be written as $\widehat{f}_{v|5}^m(\omega) = \widehat{c}_{v|5}^m(\omega) - i\widehat{q}_{v|5}^m(\omega)$, cf. theorem A.1, using a notation inspired by the one encountered when working with the complex-valued ordinary cross-spectrum. To emphasize that the present investigation is related to an auto-spectrum, the estimates $\widehat{c}_{v|5}^m(\omega)$ and $\widehat{q}_{v|5}^m(\omega)$ will respectively be referred to as estimates of the *auto-cospectrum* and the *auto-quadrature spectrum* of $\widehat{f}_{v|5}^m(\omega)$.

The bias-variance properties of $\widehat{\rho}_{v|5}(h)$ will as always depend on the number of pseudo-normalised pairs that efficiently contributes to the computation, i.e. it is of importance to once more do an analysis akin to the one done for fig. 2. In this case the requirement that a pseudo-normalised pair $(\widehat{z}_{t+h}, \widehat{z}_t)$ lies in the lag h bandwidth-square is that \widehat{z}_{t+h} lies in the bandwidth-strip centered at the 10% level, whereas \widehat{z}_t lies in the bandwidth-strip centered at the 90% level. For the present point of interest, i.e. $(-1.28, 1.28)$, a total of 59 lag 1 pairs occurred in the corresponding lag 1 bandwidth-square, which is a bit smaller than the number of pairs that occurred for the tail-points discussed in remark 3.10, which respectively was 75 and 66 for the lower and upper tail.

Contrary to the situation for the two diagonal tail-points previously investigated, the number of pseudo-normalised pairs (close to $(-1.28, 1.28)$) will increase a bit as the lag h grows up to the $h = 200$ case that was discussed in remark 3.12, in particular the number grows to 70 (which happens to coincide with the number of pseudo-normalised pairs for the lower tail). This growth from 59 to 70 seems natural when it is taken into account that the lag h pairs are expected to become independent when h grows, and then it is natural that the density should be approximately the same in regions of the same size. Note that it could be a potential problem in this example that the number of lag h pseudo-normalised pairs that efficiently contributes to the computation of $\widehat{\rho}_{v|5}(h)$ (for $h = 1, \dots, 10$) might be a bit too low, but that will be ignored here.

Figure 13 shows plots that compares the apARCH(2, 3)-model (left) with the dmbp-data (right).¹⁹ The corresponding estimated auto-cospectra $\widehat{c}_{v|5}^m(\omega)$ are given at the top and the estimated auto-quadrature spectra $\widehat{q}_{v|5}^m(\omega)$ at the bottom, with some additional details added to enable a comparison against the estimate $f^m(\omega)$ of the ordinary global m -truncated spectral density. Keeping in mind that $f^m(\omega)$ is real valued, it has only been added to the part investigating $\widehat{c}_{v|5}^m(\omega)$, and only a dotted horizontal line at $y = 0$ has been added to the plot that investigates $\widehat{q}_{v|5}^m(\omega)$.

Remark 3.20. For this particular point v and this particular truncation-level m , the estimated auto-cospectra $\widehat{c}_{v|5}^m(\omega)$ seems to indicate that the result based on independent simulations from the apARCH(2, 3)-model agrees quite well with those based on the dmbp-data. The estimated auto-quadrature spectra $\widehat{q}_{v|5}^m(\omega)$ could however indicate that there might be features of the dmbp-data that this particular GARCH-type model did not manage to pick up. Note that the aforementioned issues regarding the selection of the blocklength (to be used in the bootstrap), could imply that the pointwise confidence bands for the dmbp-data are a bit off the mark, and it would thus be premature to accept/reject a fitted model solely based on this plot alone.

¹⁹The investigation for both of the cases use the same input parameters as described earlier. In particular: 90% pointwise confidence intervals based on a 100 simulated samples of length 1974 for the apARCH-model, and similarly 90% pointwise confidence intervals based on a 100 bootstrapped replicates of length 1974 for dmbp. Both cases with the bandwidth $\mathbf{b} = (0.5, 0.5)$.

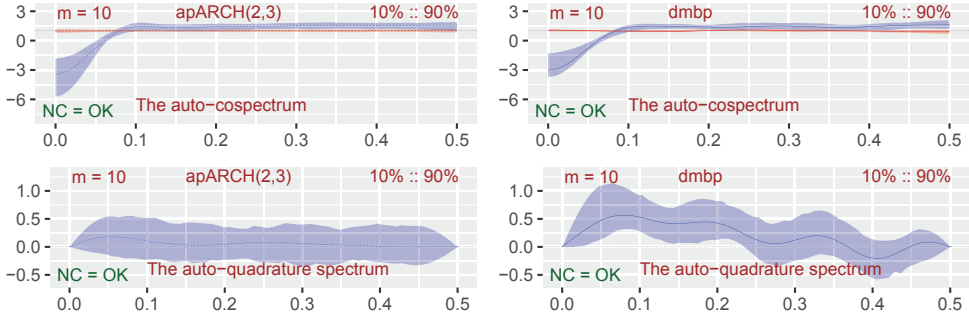


Figure 13: An off-diagonal point, comparison apARCH(2, 3)-model (left) and dmbp (right), auto-cospectrum (top) vs. auto-quadrature spectrum (bottom).

Remark 3.21. An interpretation of $\hat{c}_{v|5}^m(\omega)$ and $\hat{q}_{v|5}^m(\omega)$ from fig. 13 is beyond the aim of the present paper, but a few minor observations can nevertheless be given. First off all, observe that $\hat{q}_{v|5}^m(\omega)$ always must be 0 when $\omega \in \{0, \frac{1}{2}\}$, and then observe that a peak or (like in this case) a trough at the frequency $\omega = 0$ for $\hat{c}_{v|5}^m(\omega)$ can reveal the presence of an initial long-range sequence of same-sign estimated local Gaussian autocorrelations $\hat{\rho}_{v|5}(h)$. Furthermore, from the definitions it follows that $\hat{c}_{v|5}^m(\omega)$ and $\hat{q}_{v|5}^m(\omega)$ respectively must integrate to 1 and 0 over $(-\frac{1}{2}, \frac{1}{2}]$, and those properties might be useful when discussing peaks/troughs that occurs in the graphs.

Remark 3.22. When lemma 2.3(c) is taken into account, it is clear that the plot of $\hat{q}_{v|5}^m(\omega)$ can give an indication of whether or not the time series under investigation is reversible. For the GARCH-type model investigated in fig. 13, the $\hat{q}_{v|5}^m(\omega)$ -plot indicates that it might represents a reversible time series, whereas the dmbp-data on the other and does not quite give that impression. However, it would be premature to draw any conclusions based on the particular combination of point v , truncation level m and bandwidth b used in this example.

Remark 3.23. A single plot based on the amplitude of $\hat{f}_{v|5}^m(\omega)$ could have revealed approximately the same as the present plots of $\hat{c}_{v|5}^m(\omega)$, but it would not necessarily reveal the details that $\hat{q}_{v|5}^m(\omega)$ can give about the reversibility of the time series under investigation. Moreover, a peak at $\omega = 0$ of the amplitude-specter would detect the presence of some long-range dependency of $\hat{\rho}_{v|5}(h)$, but it might not be immediately clear from the amplitude-plot whether the prevailing sign of these $\hat{\rho}_{v|5}(h)$ was positive or negative.

Remark 3.24. This example shows that $\hat{f}_{v|5}^m(\omega)$ might provide some insight also when the point v lies off the diagonal. In a practical setting, it would of course be natural/necessary to investigate several different off-diagonal points v for different combinations of bandwidths b and truncation levels m , and for plots based on real data it would also be preferable to test different blocklengths to see how much the estimated pointwise confidence bands depends on that setting.

3.7 1 parameter versus 5 parameter

The plots considered so far have all used the 5 parameter local Gaussian approximation in the computation of the local Gaussian autocorrelations, i.e. estimates $\hat{\rho}_{v|5}(h)$ of $\rho_{v|5}(h)$ have been used in the computations. The reason for this is that $\hat{\rho}_{v|1}(h)$ in general might not carry

sufficient information about the local correlation structure of the densities, cf. the discussion in appendix C.6. The estimates $\hat{f}_{v|1}^m(\omega)$ might thus return rather dubious results.

The inadequacy of $\hat{f}_{v|1}^m(\omega)$ is highlighted in fig. 14 where it has been compared against $\hat{f}_{v|5}^m(\omega)$ for the previously encountered comparison of the GARCH-model and the dmbp-data (all the parameters are identical to those described earlier). The two upper panels show the previously encountered results for the simulated GARCH-model (left side) and the dmbp-data (right side), whereas the two lower panels shows the corresponding results when $\hat{\rho}_{v|1}(h)$ was used instead of $\hat{\rho}_{v|5}(h)$. The differences that occurs for the lower frequencies in these cases are quite clear, i.e. the results obtained from the 1 parameter approach are not as high as those obtained from the 5 parameter approach, and it might thus happen that an analysis based on $\hat{f}_{v|1}^m(\omega)$ does not detect a difference that is clear-cut when $\hat{f}_{v|5}^m(\omega)$ is employed. This makes the $\hat{f}_{v|1}^m(\omega)$ less favourable as an exploratory tool.²⁰

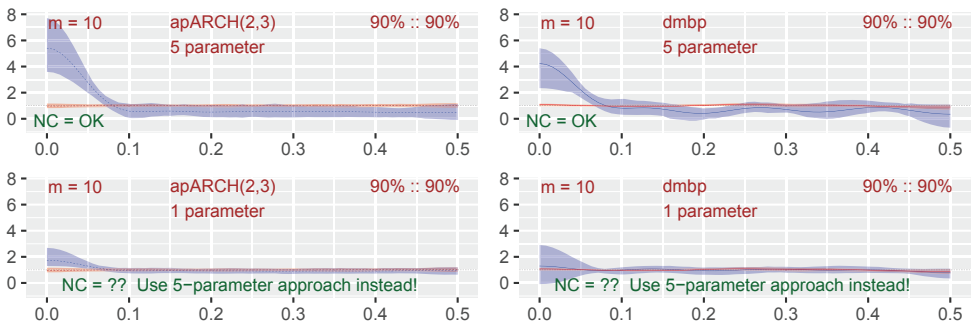


Figure 14: apARCH(2,3)-model (left) and dmbp (right), 5 parameter (top) vs. 1 parameter (bottom).

The ‘trumpet shaped’ pointwise confidence band observed in the lower right panel seems to be a common feature when the 1 parameter approach $\hat{f}_{v|1}^m(\omega)$ are used on non-Gaussian data (like the dmbp-example encountered here). This phenomenon occurs for a wide range of different blocklengths for the bootstrap, which implies that the ‘problem’ is not directly related to the blocklength that was used in this particular case. Due to larger flexibility, it seems evident that the 5 parameter approach $\hat{f}_{v|5}^m(\omega)$ is the better option to apply. See the discussion in appendix C.6 for further details.

For the record, these undesirable differences does not occur for Gaussian time series, which is as expected since both the 1 parameter and 5 parameter algorithms in such nice cases will give an estimated local Gaussian autocorrelation close to the ordinary autocorrelation of the Gaussian time series. But, as the cases considered in fig. 14 shows, it seems clear that the 1 parameter approach $\hat{f}_{v|1}^m(\omega)$ might not be up to the task when the structure of interest deviates from the Gaussian assumption.

²⁰Note that the algorithm used to find $\hat{\rho}_{v|1}(h)$ does not reveal whether or not the result is based on a successful numerical computation, and this is the reason that ‘NC=??’ is shown in fig. 14.

4 Discussion

The examples in section 3 show that an exploratory tool based on estimates of the local Gaussian spectral density $f_{v|5}(\omega)$ might be useful, and that it in some cases might be possible to interpret peaks and troughs that occurs in a manner similar to the interpretation used when estimates of the ordinary spectral density $f(\omega)$ are inspected. Caution must however be exercised, since there still are many details related to the selection of the point v and the parameters m and b that needs to be investigated further, and this section will present some additional comments related to this part. Moreover, the unresolved issue with regard to the selection of the blocklength (when bootstrapping is needed) will also be discussed here, before an alternative smoothing strategy is commented upon at the end of this section.

4.1 The points v and the bandwidths b

Based on the discussion in section 3.2, it is clear that the combination of a point v and a bandwidth b has a large impact on the number of pseudo-normalised lag h pairs from $\{(\hat{z}_{t+h}, \hat{z}_t)\}_{t=1}^{n-h}$ that efficiently contributes to the computation of the estimates $\hat{\rho}_{v|5}(h)$.

The selection of v and b must be seen in conjunction, see remark 3.10, and in particular: If b is rather small, then it is important to not select points v too far from the center of the distribution, since the small-sample variation then might become the dominating feature of the estimated values $\hat{f}_{v|5}^m(\omega)$.

In section 3, the bandwidth $b = (.5, .5)$ was used for the majority of the plots, with the exception of fig. 6, page 23. Figure 6 indicates that the estimates of $f^m(\omega)$ and $f_{v|5}^m(\omega)$ (the red and blue dashed lines) seems to be the same for the three bandwidths chosen, but the width of the corresponding pointwise confidence bands changes quite a bit, i.e. they, as expected, become narrower with higher bandwidths.

The bandwidth $b = (.5, .5)$ used as default in section 3 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 experimentation with the bandwidth-selection algorithm used in Otneim and Tjøstheim (2016), and it has been applied here even though it originates from a bandwidth-selection algorithm aimed at computing density estimates based on the one-parameter local Gaussian approximation.

It might be a dubious practice to use the same bandwidth for all the lags $h = 1, \dots, m$, and it could also be a problem that the same bandwidth is used for all the points v , since the number of observations in the vicinity of points in the tail is much smaller than the corresponding number for a point in the center, cf. remark 3.10. However, used as an exploratory tool, with pointwise confidence bands that clearly shows the different variances, it should still make sense to use the same bandwidth for a comparison like the one between the apARCH(2, 3)-model from fig. 10 and the original dmbp-data in fig. 11 (see page 29).

Remark 4.1. The R code used for the estimation of the local Gaussian autocorrelations, i.e. $\rho_{v|5}(h)$ and $\rho_{v|1}(h)$, can apply different alternatives for the bandwidth-argument. It is e.g. possible to use an approach where a percentage is given, and the algorithm then selects for each point v and each lag h a bandwidth that ensures that this percentage of the available pseudo-normalised lag h pairs are included in the resulting bandwidth-square. A few experiments with this simplistic bandwidth-approach did not produce results that differed significantly from those based on fixed bandwidths.

Remark 4.2. A simple *rule of thumb*, like the formula $b \approx 1.75n^{-1/6}$, would be preferable with regard to the selection of the bandwidth, since the computational cost can become a problem if a new bandwidth has to be computed for each of the lags $h = 1, \dots, m$. In particular, if a selection-algorithm for \mathbf{b} could be based on a cross-validation technique, then it would for practical purposes probably be preferable to first compute estimates of $f_{v|5}^m(\omega)$ for a range of fixed bandwidths, and then apply the selection-algorithm only for those cases where some potentially interesting structures was revealed in the initial investigation.

Remark 4.3. 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. An earlier version of the R code (used for the estimations in the present paper) had an option where the $p = 1$ version of this algorithm, from Otneim and Tjøstheim (2016, Section 4), could be used when estimating $\rho_{v|1}(h)$. This option did however result in a tremendous increase in the computational time, since the estimation of $f_{v|1}^m(\omega)$ requires the estimation of m different local Gaussian auto-correlations $\rho_{v|1}(h)$.²¹ Moreover, it is also a 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_v -out cross-validation, h -block cross validation, and hv -block cross-validation were introduced as better tools for the dependent case.

4.2 The truncation level m and the weighting function $\lambda_m(h)$

For estimates of the ordinary spectral density, $f(\omega)$, there exist rules of thumb (based on the number of observations n) that can define a range within which an appropriate truncation level m might be found, such that a reasonable bias-variance balance is obtained for the estimated spectral density $\hat{f}(\omega)$. The guiding principle for the selection of m for the global case is based on the observation that there is a linear decrease in the number of lag h pairs, so the variance of the estimates $\hat{\rho}(h)$ increases for higher lag-values, and the selection of truncation level m and weighting function $\lambda_m(h)$ is then used to counter the effect of this increased variance from high-lag components.

It would be preferable to have some similar guiding principle for the selection of m for the local case too, but in this case the situation is more complicated since the bias-variance properties of the building blocks $\hat{\rho}_{v|5}(h)$ are affected both by the position of the point \mathbf{v} and the selected bandwidth \mathbf{b} . In particular, the kernel function involved in the estimation of $\rho_{v|5}(h)$ implies that the variance will depend on the number of pseudo-normalised lag h pairs $(\hat{z}_{t+h}, \hat{z}_t)$ that lies inside the lag h bandwidth-squares, as shown in fig. 2.

Remarks 3.10 and 3.12 describe (for the three investigated diagonal points) the number of efficiently contributing pseudo-normalised lag h pairs for the two lags $h = 1$ and $h = 200$. If a common truncation level m is to be used for all the three points, then it is clear that the points \mathbf{v} with the smallest number of contributing pseudo-normalised pairs should be considered, i.e.

²¹Tests 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 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.

the numbers for the points at the lower and upper tails are those that is central to the selection of truncation level m .

For these tail-points, the reduction from the lag 1 case to the lag 200 case was rather small, which could imply that the variance of the corresponding estimates $\widehat{\rho}_{v|5}(h)$ grows at a much slower pace than the variance of the estimates of the global autocorrelation $\widehat{\rho}(h)$. Furthermore, as the off-diagonal example in section 3.6 shows, points v can be found where the number of efficiently contributing lag h pairs increases when h grows from lag 1 to lag 200.

In lack of a data-driven rule that can propose a suitable range of values for the truncation level m , the following strategy has been used instead: First estimate the local Gaussian autocorrelations $\widehat{\rho}_{v|5}(h)$ for a large range of lags, and then use a shiny-application (see footnote 17, page 18) to interactively play through the plots of the corresponding m truncated estimates $\widehat{f}_{v|5}^m(h)$. A drawback with this solution is that it might waste computational resources on cases where small-sample variation distorts the presence of any local signals.

4.3 The blocklength for the bootstrap

There do exist data-driven methods for the selection of the blocklength to be used when bootstrapping, 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 does not give a good result 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 blocklength algorithms (both for the stationary and for the circular bootstrap) considers a block of length 1 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, there does not exist an adjustment of the blocklength algorithm suited for the present case of interest. This implies that the use of the local Gaussian spectral density on real data suffers from the problem that the blocklength for the bootstrap must be manually selected, which makes it harder to decide if a potentially interesting difference between the ordinary and local spectral density really should be considered to be significant – or if it should be discarded as a spurious effect due to a badly selected blocklength for the construction of the pointwise confidence bands.

As explained in remark 3.19: The blocklength 100 was used for the dmbp-example (see fig. 11) in order to get plots where the estimate $\widehat{f}_{v|5}^m(\omega)$ based on the original sample was positioned approximately at the center of the resulting pointwise confidence-band.²² An approach based on the testing of several different blocklengths is computationally costly, so it would be preferable to find some data-driven strategy.

Based on the selection-algorithm in (Politis and White, 2004), one might wonder if an adjusted selection algorithm suited for the local case could be created by replacing the estimated

²²It should be noted that a wide range of possible blocklengths was investigated, and they all revealed significant differences between the ordinary and local spectrum for low frequencies in the lower and upper tails – so the dmbp-investigation did most likely detect an actual phenomenon in the data at hand.

autocovariances $R(h)$ with local Gaussian autocorrelations $\widehat{\rho}_{v|5}(h)$ instead. A potential problem with this approach is that the result could depend upon the points v that are investigated. But still, if nothing else, a visual inspection (like the one given in fig. 3) of the estimated values $\widehat{\rho}_{v|5}(h)$ might help motivate lower and upper thresholds within which a search for the blocklength could be restricted. From this a blocklength of 100 might not be unreasonable.

It also has to be noted that there is an additional issue that remains to be investigated, and that is the asymptotic properties of the bootstrap-approach in this particular case. The theoretical properties of the bootstrap-methodology in the realm of local Gaussian correlation have been investigated in Lacal and Tjøstheim (2017), and it seems likely that the methods employed there could be adjusted to cover the present case of interest.

Remark 4.4. As mentioned before, the R-package `localgaussSpec` allows the estimation of $\widehat{f}_{v|5}^m(\omega)$ for a wide range of settings for the points and parameters, with a simple interactive solution for the inspection of the results. It is thus not that critical that data-driven methods for the selection of the parameters still are missing, but with regard to the computational costs it would be preferable to have at least some guiding principles that could restrict the initial attention to parameter-regions where small-sample variance should not distort the presence of any local signals.

4.4 An alternative smoothing strategy?

The previously defined estimates $\widehat{f}_{v|5}^m(\omega)$ of $f_{v|5}(\omega)$ was based on a weighting function $\lambda_m(h)$ that worked upon the estimated values $\widehat{\rho}_{v|5}(h)$, but it should for the record be noted that an alternative approach could have been applied too.

The point is, as mentioned in remark B.4, that it is possible to extend the result of appendix B.4 to show that the estimated m -truncated local Gaussian spectral densities $\widehat{f}_{v_i|5}^m(\omega)$ corresponding to different points $\{v_i\}_{i=1}^v$ will be jointly asymptotically normal and pairwise asymptotically independent (when $m \rightarrow \infty$ and $\mathbf{b} \rightarrow \mathbf{0}^+$ as $n \rightarrow \infty$). This enables an alternative smoothing strategy, where an estimate $\widehat{f}_{v|5}^m(\omega)$ for a given point v could be based on a weighting of the values of $\widehat{f}_{v_i|5}^m(\omega)$ in a grid of points surrounding 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 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.

5 Summary

This paper presents *the local Gaussian spectral density* $f_{v|p}(\omega)$ as a new possible approach to the study of nonlinear time-series and nonlinear periodic phenomena. This method is based upon the simple approach that the ordinary autocorrelations $\rho(h)$ in the standardised expression for the spectral density, see eq. (1.2), are replaced with the local Gaussian autocorrelation $\rho_{v|p}(h)$.

Estimates $\hat{f}_{v|p}^m(\omega)$ of the (m truncated) local Gaussian spectral densities are then obtained by estimating the corresponding local Gaussian autocorrelations.

The parameter p can either be 5 or 1 depending on the type of the number of parameters used in the local Gaussian approximation. The theory covers both options, but for practical purposes, based on our experiences, only $p = 5$ can be recommended, since the $p = 1$ case fails to capture the local structures in a satisfying way, see fig. 14 on page 33 for an illustration.

The point $\boldsymbol{v} = (v_1, v_2)$ can in principle be any point in \mathbb{R}^2 , but it is important to keep in mind that small-sample variation could become a problem if it lies in an area with few observations. Moreover, the visualisation and interpretation of the results might be easier if \boldsymbol{v} lies on the diagonal, i.e. $v_1 = v_2$, since (as seen in theorem 2.7) real-valued results are obtained in that case.

The examples²³ in section 3 indicates that this method can be used as an exploratory tool to detect the presence of local structures that the ordinary spectral density does not register, and that it also could give some aid when it comes to selecting models fitted to data, cf. the discussion relating the apARCH(2, 3)-model from section 3.3 with the dmbp-data in section 3.4. It seems likely that such a comparison could be of interest even when it is not known whether or not the investigated time series satisfies the requirements needed for the asymptotic theory to work.

Finally, it should be noted that this paper only aims at presenting the method and that there are many issues that remains to be resolved with regard to the use of this method. This includes, as discussed in section 4.1, the need for some rules of thumbs with regard to how far out in the tails it makes sense to select the points \boldsymbol{v} given a number n of observations, the need for some suitable method to select the bandwidths \boldsymbol{b} and the truncation point m for the given number of observations (potentially also depending on the point \boldsymbol{v}) – and of course the issue regarding the blocklength to use when working upon non-linear white noise. For all of these arguments there is a need for a better understanding of the effect of them upon the bias-variance balance of the resulting estimates.

²³The scripts for all the examples are included in the R-package `localgaussSpec`, available on github

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

This appendix presents the asymptotic properties of $\widehat{f}_{v|p}^m(\omega)$, the m -truncated estimate of the local Gaussian spectral density, i.e. the proof of theorem 2.7 is given here together with a theorem that covers the case when $\widehat{f}_{v|p}^m(\omega)$ is complex-valued. The technical details needed for the proofs are covered in appendices B and C

A.1 The proof of theorem 2.7

Proof. The property that $f_{v|p}(\omega)$ is a real-valued function when v lies on the diagonal was proved in lemma 2.3(d). The expression for $\widehat{f}_{v|p}^m(\omega)$ from lemma 2.5 can by the help of vectors be written as

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

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

$$\Lambda'_m(\omega) := [\lambda_m(1) \cdot \cos(2\pi\omega \cdot 1), \dots, \lambda_m(m) \cdot \cos(2\pi\omega \cdot m)], \quad (\text{A.2a})$$

$$\widehat{P}_{v|m|b|p} := [\widehat{\rho}_{v|p}(1|b_1), \dots, \widehat{\rho}_{v|p}(m|b_m)]'. \quad (\text{A.2b})$$

Since $\widehat{\rho}_{v|p}(h|b_h)$ is one of the p estimated parameters $\widehat{\theta}_{v|p}(h|b_h)$ from the local Gaussian approximation (of the lag h pairs) at the point v ,²⁴ it is clear that it is possible to write $\widehat{\rho}_{v|p}(h|b_h) = e'_p \cdot \widehat{\theta}_{v|p}(h|b_h)$, where e'_p is the unit vector that picks out $\widehat{\rho}_{v|p}(h|b_h)$ from $\widehat{\theta}_{v|p}(h|b_h)$. The vectors $\{\widehat{\theta}_{v|p}(h|b_h)\}_{h=1}^m$ can be stacked on top of each other to give a joint parameter vector $\widehat{\theta}_{v|\overline{m}|b|p}$, and it follows that the vector $\widehat{P}_{v|m|b|p}$ can be expressed as $\widehat{P}_{v|m|b|p} = \mathbf{E}'_{m|p} \cdot \widehat{\theta}_{v|\overline{m}|b|p}$, where $\mathbf{E}'_{m|p}$ is the matrix that picks out the relevant components from $\widehat{\theta}_{v|\overline{m}|b|p}$. ($\mathbf{E}'_{m|p}$ will be the $m \times m$ identity matrix if $p = 1$.) It follows from this, and Brockwell and Davis (1986, Proposition 6.4.2, p. 211), that an asymptotic normality result for $\widehat{\theta}_{v|\overline{m}|b|p}$ will give an asymptotic normality result for $\widehat{f}_{v|p}^m(\omega)$. In particular, if a suitable scaling factor²⁵ $c_{n|m|b|p}$ gives an mp -variate asymptotic normality result for $\widehat{\theta}_{v|\overline{m}|b|p}$,

$$c_{n|m|b|p} \cdot \left(\widehat{\theta}_{v|\overline{m}|b|p} - \theta_{v|\overline{m}|p} \right) \xrightarrow{d} \mathbf{N}(\mathbf{0}, \Sigma_{v|\overline{m}|p}), \quad (\text{A.3})$$

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

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

²⁴The properties of $\widehat{\theta}_{v|p}(h|b_h)$ 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.

²⁵ $c_{n|m|b|p}$ must be a function of n , m and $\{b_h\}_{h=1}^m$, such that $c_{n|m|b|p} \rightarrow \infty$ when $n \rightarrow \infty$, $m \rightarrow \infty$ and $b_h \rightarrow \mathbf{0}^+$.

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

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

The asymptotic normality required in eq. (A.3) follows from theorem B.22 (page 64), i.e. the scaling factor $c_{n|m|b|p}$ will be $\sqrt{n(b_1 b_2)^{(p+1)/2}}$, whereas the asymptotic covariance matrix $\Sigma_{\mathbf{v}|\overline{m}|p}$ can be written as the direct sum of the covariance matrices for $\sqrt{n(b_1 b_2)^{(p+1)/2}} \cdot \widehat{\boldsymbol{\theta}}_{\mathbf{v}|p}(h|b_h)$, i.e.

$$\text{Var}\left(\sqrt{n(b_1 b_2)^{(p+1)/2}} \cdot \widehat{\boldsymbol{\theta}}_{\mathbf{v}|\overline{m}|b|p}\right) = \bigoplus_{h=1}^m \text{Var}\left(\sqrt{n(b_1 b_2)^{(p+1)/2}} \cdot \widehat{\boldsymbol{\theta}}_{\mathbf{v}|p}(h|b_h)\right), \quad (\text{A.6})$$

from which a simple calculation gives

$$\text{Var}\left(\sqrt{n(b_1 b_2)^{(p+1)/2}} \cdot \widehat{f}_{\mathbf{v}|p}^m(\omega)\right) = 4 \cdot \sum_{h=1}^m \lambda_m^2(h) \cdot \cos^2(2\pi\omega h) \cdot \text{Var}\left(\sqrt{n(b_1 b_2)^{(p+1)/2}} \cdot \widehat{\rho}_{\mathbf{v}|p}(h|b_h)\right). \quad (\text{A.7})$$

From this it is clear that the scaling factor $c_{n|m|b|p}$ 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, $c'_{n|m|b|p} = \sqrt{n(b_1 b_2)^{(p+1)/2}/m}$, which completes the proof. \square

Remark A.1. Some care must be taken formally with regard to the limiting mp -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 $f_{\mathbf{v}|p}(\omega)$ is a complex valued function for a point $\mathbf{v} = (v_1, v_2)$, i.e. $f_{\mathbf{v}|p}(\omega) = c_{\mathbf{v}|p}(\omega) - iq_{\mathbf{v}|p}(\omega)$, with $q_{\mathbf{v}|p}(\omega) \not\equiv 0$, then, under assumptions 2.1 to 2.3, the components $\widehat{c}_{\mathbf{v}|p}^m(\omega)$ and $\widehat{q}_{\mathbf{v}|p}^m(\omega)$ of the m -truncated estimate $\widehat{f}_{\mathbf{v}|p}^m(\omega)$ will, when $\omega \notin \frac{1}{2} \cdot \mathbb{Z} := \{\dots, -1, -\frac{1}{2}, 0, \frac{1}{2}, 1, \dots\}$, be jointly asymptotically normally distributed as given below.*

$$\sqrt{n(b_1 b_2)^{(p+1)/2}/m} \cdot \left(\begin{bmatrix} \widehat{c}_{\mathbf{v}|p}^m(\omega) \\ \widehat{q}_{\mathbf{v}|p}^m(\omega) \end{bmatrix} - \begin{bmatrix} c_{\mathbf{v}|p}(\omega) \\ q_{\mathbf{v}|p}(\omega) \end{bmatrix} \right) \xrightarrow{d} \mathbf{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{c;\mathbf{v}|p}^2(\omega) & 0 \\ 0 & \sigma_{q;\mathbf{v}|p}^2(\omega) \end{bmatrix} \right), \quad (\text{A.8})$$

where the variances $\sigma_{c:v|p}^2(\omega)$ and $\sigma_{q:v|p}^2(\omega)$ are given by

$$\sigma_{c:v|p}^2(\omega) = \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{h=1}^m \lambda_m^2(h) \cdot \cos^2(2\pi\omega h) \cdot \{\tilde{\sigma}_{v|p}^2(h) + \tilde{\sigma}_{\check{v}|p}^2(h)\} \quad (\text{A.9a})$$

$$\sigma_{q:v|p}^2(\omega) = \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{h=1}^m \lambda_m^2(h) \cdot \sin^2(2\pi\omega h) \cdot \{\tilde{\sigma}_{v|p}^2(h) + \tilde{\sigma}_{\check{v}|p}^2(h)\}, \quad (\text{A.9b})$$

with $\tilde{\sigma}_{v|p}^2(h)$ and $\tilde{\sigma}_{\check{v}|p}^2(h)$ related to respectively $\hat{\rho}_{v|p}(h|b_1)$ and $\hat{\rho}_{\check{v}|p}(h|b_1)$ as given in theorem 2.7.

The component $\hat{q}_{v|p}^m(\omega)$ 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

$$\sqrt{n(b_1 b_2)^{(p+1)/2} / m} \cdot \left(\hat{f}_{v|p}^m(\omega) - f_{v|p}(\omega) \right) \xrightarrow{d} \mathbf{N}(0, \sigma_{c:v|p}^2(\omega)), \quad (\text{A.10})$$

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 2.7, whereas the general case requires a bivariate extension of that proof. In particular, when the proof of theorem 2.7 is used on $\hat{c}_{v|p}^m(\omega)$ and $\hat{q}_{v|p}^m(\omega)$, it follows that they can be written as

$$\hat{c}_{v|p}^m(\omega) = 1 + \Lambda'_{c|m}(\omega) \cdot \hat{\mathbf{P}}_{v|m|b|p} + \Lambda'_{c|m}(\omega) \cdot \hat{\mathbf{P}}_{\check{v}|m|b|p} = 1 + \Lambda'_{c|\bar{m}}(\omega) \cdot \hat{\mathbf{P}}_{v|\bar{m}|b|p} \quad (\text{A.11a})$$

$$\hat{q}_{v|p}^m(\omega) = 0 + \Lambda'_{q|m}(\omega) \cdot \hat{\mathbf{P}}_{v|m|b|p} - \Lambda'_{q|m}(\omega) \cdot \hat{\mathbf{P}}_{\check{v}|m|b|p} = 0 + \Lambda'_{q|\bar{m}}(\omega) \cdot \hat{\mathbf{P}}_{v|\bar{m}|b|p}, \quad (\text{A.11b})$$

where $\Lambda'_{c|m}(\omega)$ and $\Lambda'_{q|m}(\omega)$ are the coefficient vectors containing respectively the cosines and sines, where $\hat{\mathbf{P}}_{v|m|b|p}$ and $\hat{\mathbf{P}}_{\check{v}|m|b|p}$ contains the estimated correlations corresponding to v and \check{v} for the lags under consideration, and where the length $2m$ vectors $\Lambda'_{c|\bar{m}}(\omega)$, $\Lambda'_{q|\bar{m}}(\omega)$ and $\hat{\mathbf{P}}_{v|\bar{m}|b|p}$ 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 2.7, it follows that $\hat{\mathbf{P}}_{v|\bar{m}|b|p} = (\mathbf{E}'_{m|p} \oplus \mathbf{E}'_{m|p}) \cdot \hat{\Theta}_{\bar{m}|b|p}(v, \check{v})$, where $\hat{\Theta}_{\bar{m}|b|p}(v, \check{v})$ is the full set of estimated parameters from the local Gaussian approximations at v and \check{v} for the lags under consideration,²⁶ and where $(\mathbf{E}'_{m|p} \oplus \mathbf{E}'_{m|p})$ is the matrix that picks out the relevant autocorrelations.

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

$$\begin{bmatrix} \hat{c}_{v|p}^m(\omega) \\ \hat{q}_{v|p}^m(\omega) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} \Lambda'_{c|\bar{m}}(\omega) \\ \Lambda'_{q|\bar{m}}(\omega) \end{bmatrix} \cdot (\mathbf{E}'_{m|p} \oplus \mathbf{E}'_{m|p}) \cdot \hat{\Theta}_{\bar{m}|b|p}(v, \check{v}), \quad (\text{A.12})$$

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

$$\sqrt{n(b_1 b_2)^{(p+1)/2}} \cdot \left(\hat{\Theta}_{\bar{m}|b|p}(v, \check{v}) - \Theta_{\bar{m}|b|p}(v, \check{v}) \right) \xrightarrow{d} \mathbf{N}(0, \Sigma_{v|\bar{m}|p} \oplus \Sigma_{\check{v}|\bar{m}|p}), \quad (\text{A.13})$$

gives the result when the arguments in the proof of theorem 2.7 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 $\sigma_{q:v|p}^2(\omega)$ 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. \square

²⁶The vector $\hat{\Theta}_{\bar{m}|b|p}(v, \check{v})$ can be expressed as a combination of $\hat{\theta}_{v|\bar{m}|b|p}$ and $\hat{\theta}_{\check{v}|\bar{m}|b|p}$, where $\hat{\theta}_{v|\bar{m}|b|p}$ is the parameter vector from the proof of theorem 2.7.

Appendix B: Asymptotic results for $\hat{\theta}_{v|\bar{m}|b|p}$

This section will investigate the asymptotic properties of the parameter vector $\hat{\theta}_{v|\bar{m}|b|p}$, that is used in the proof of theorem 2.7. The proof is similar in spirit to the one used in Tjøstheim and Hufthammer (2013) for the asymptotic investigation of the parameter vectors $\hat{\theta}_{v|p}(h|b_h)$, 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 $\hat{\theta}_{v|\bar{m}|b|p}$ are collected in appendix B.3.

Remark B.1. The asymptotic investigation requires several indices in order to keep track of the different components, and to simplify references to v , b and p 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* $\rho_{v|5}$ as a new *local measure of dependence* at a point v , and then used the approach formalised in Klimko and Nelson (1978), to investigate the asymptotic properties of $\hat{\rho}_{v|5}$. (The arguments in (Tjøstheim and Hufthammer, 2013) holds for the simpler case $p = 1$ too.) The *local Gaussian spectral density* $f_{v|p}(\omega)$ is based on the local Gaussian autocorrelations $\rho_{v|p}(h)$, and the asymptotic properties of the estimates $\hat{f}_{v|p}^m(\omega)$ are thus closely connected to the asymptotic properties of $\hat{\rho}_{v|p}(h)$.

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

such that the following Taylor expansion is valid (in the neighbourhood \mathcal{N}) for $\|\boldsymbol{\theta} - \boldsymbol{\theta}^\circ\| < \delta$,

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

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

where V_n and $T_n(\boldsymbol{\theta}^*)$ are defined in the obvious manner, with $\boldsymbol{\theta}^* = \boldsymbol{\theta}^*(\mathbf{X}_1, \dots, \mathbf{X}_n; \boldsymbol{\theta})$ 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 $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ and Q_n are such that as $n \rightarrow \infty$*

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

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

$$(A3) \quad \text{for } j, k = 1, \dots, p$$

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

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

Then there exists a sequence of estimators $\widehat{\boldsymbol{\theta}}_n = (\widehat{\theta}_1, \dots, \widehat{\theta}_p)'$, 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° 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

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

then

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

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 $\psi(\mathbf{y}_h; \boldsymbol{\theta}_{v|h})$ to approximate the bivariate densities $g_h(\mathbf{y}_h)$ at a point \mathbf{v} , where $\boldsymbol{\theta}_{v|h} = [\theta_{v|h:1}, \dots, \theta_{v|h:5}]'$ is the five dimensional parameter-vector of the bivariate Gaussian distribution. The point \mathbf{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 $\boldsymbol{\theta}_{v|h}$.

The local investigation requires a bandwidth vector $\mathbf{b} = (b_1, b_2)$ and a kernel function $K(\mathbf{w})$, which is used to define $K_{h;\mathbf{b}}(\mathbf{y}_h - \mathbf{v}) := \frac{1}{b_1 b_2} K\left(\frac{y_h - v_1}{b_1}, \frac{y_0 - v_2}{b_2}\right)$, which in turn is used in the following local approximation around \mathbf{v} ,

$$q_{n;\mathbf{b}} := \int_{\mathbb{R}^2} K_{h;\mathbf{b}}(\mathbf{y}_h - \mathbf{v}) [\psi(\mathbf{y}_h; \boldsymbol{\theta}_h) - g_h(\mathbf{y}_h) \log \psi(\mathbf{y}_h; \boldsymbol{\theta}_h)] d\mathbf{y}_h, \quad (\text{B.4})$$

a minimiser of which should satisfy the vector equation

$$\int_{\mathbb{R}^2} K_{h;b}(\mathbf{y}_h - \mathbf{v}) \mathbf{u}_h(\mathbf{y}_h; \boldsymbol{\theta}_h) [\psi(\mathbf{y}_h; \boldsymbol{\theta}_h) - g_h(\mathbf{y}_h)] d\mathbf{y}_h = \mathbf{0}, \quad (\text{B.5})$$

where $\mathbf{u}_h(\mathbf{y}_h; \boldsymbol{\theta}_h) := \nabla_h \log \psi(\mathbf{y}_h; \boldsymbol{\theta}_h)$ is the score function of $\psi(\mathbf{y}_h; \boldsymbol{\theta}_h)$ (with $\nabla_h := \partial/\partial\boldsymbol{\theta}_h$). Under the assumption that there is a bandwidth \mathbf{b}_0 such that there exists a minimiser $\boldsymbol{\theta}_{h;b}$ of eq. (B.4) which satisfies eq. (B.5) for any \mathbf{b} with $\mathbf{0} < \mathbf{b} < \mathbf{b}_0$,²⁷ this $\boldsymbol{\theta}_{h;b}$ will be referred to as the population value for the given bandwidth \mathbf{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.

For a sample of size n from $\{\mathbf{Y}_{h:t}\}_{t \in \mathbb{Z}^*}$, the following M -estimator²⁸ will be used, which (due to the ergodicity implied by assumption 2.1(a)) will converge towards the penalty function $q_{h;b}$,

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

The local penalty function from (Tjøstheim and Hufthammer, 2013) can be described as

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

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 help of the score function $\mathbf{u}_h(\mathbf{y}_h; \boldsymbol{\theta}_h)$ can be given as,

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

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 $\nabla_h Q_{h;n}(\boldsymbol{\theta}_h)$ is evaluated at the population value $\boldsymbol{\theta}_{h;b}$.

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

²⁷Inequalities involving vectors are to be interpreted in a component-wise manner.

²⁸The 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

of $\boldsymbol{\theta}_{h:b}$, i.e.

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

where

$$V_{h:b:n} := V_{h:b:n}(\boldsymbol{\theta}_{h:b}) := \nabla_h \nabla_h' Q_{h:n}(\boldsymbol{\theta}_{h:b}), \quad (\text{B.9b})$$

$$T_{h:b:n} := T_{h:b:n}(\boldsymbol{\theta}_h^*, \boldsymbol{\theta}_{h:b}) := \nabla_h \nabla_h' Q_{h:n}(\boldsymbol{\theta}_h^*) - \nabla_h \nabla_h' Q_{h:n}(\boldsymbol{\theta}_{h:b}), \quad (\text{B.9c})$$

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

With the preceding definitions, (Tjøstheim and Hufthammer, 2013, theorem 1) investigated the case where the bandwidth \mathbf{b} was fixed as $n \rightarrow \infty$, 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 \mathbf{v}) with $P(A_h^c) < \epsilon$, such that there exists a sequence of estimators $\widehat{\boldsymbol{\theta}}_{h:n}$ that converges almost surely to $\boldsymbol{\theta}_{h:b}$ (the minimiser of $q_{h:b}$ from eq. (B.4)). And, moreover, the following asymptotic behaviour is observed

$$(nb_1 b_2)^{1/2} \left(\widehat{\boldsymbol{\theta}}_{h:n} - \boldsymbol{\theta}_{h:b} \right) \xrightarrow{d} \text{N}(\mathbf{0}, \Sigma_{h:b}), \quad (\text{B.10})$$

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

The situation when $\mathbf{b} \rightarrow \mathbf{0}^+$ as $n \rightarrow \infty$ requires some extra care since the presence of the kernel function $K_{h:b}(\mathbf{w})$ in $Q_{h:n}(\boldsymbol{\theta}_h)$, see eq. (B.7), gives limiting matrices of $V_{h:b}$ and $W_{h:b}$ 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 $\mathbf{b} = (b_1, b_2)$ are such that $\log n/n(b_1 b_2)^5 \rightarrow 0$,

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

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

B.1.3 A new penalty function

The proof of theorem 2.7 requires an asymptotic result for the parameter vector $\widehat{\boldsymbol{\theta}}_{n|\bar{m}|b|p}$, which was obtained by combining m parameter vectors corresponding to the bivariate lag h pairs (Y_{t+h}, Y_t) for $h = 1, \dots, m$. This section will show (for the case $p = 5$) how a penalty function for $\widehat{\boldsymbol{\theta}}_{n|\bar{m}|b|p}$ can be constructed based on the bivariate penalty functions $Q_{h:n}$ defined in appendix B.1.2. The indices n , \mathbf{b} and p will for notational simplicity be suppressed from the notation, and only $\boldsymbol{\theta}_{\bar{m}}$ 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 \mathbf{b} both are fixed as $n \rightarrow \infty$. 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 $m \rightarrow \infty$ and $\mathbf{b} \rightarrow \mathbf{0}^+$ when $n \rightarrow \infty$, 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 $\{\mathbf{Y}_{h:t}\}_{t \in \mathbb{Z}}$, whereas the new penalty function will be defined for the $(m+1)$ -variate time series $\{\mathbf{Y}_{\bar{m}:t}\}_{t \in \mathbb{Z}}$. The first step is to extend the penalty functions $Q_{h:n}$, $h = 1, \dots, m$ from expression based on $\mathbf{Y}_{h:t}$ to expressions based on $\mathbf{Y}_{\bar{m}:t}$, 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 remark 2.9, which gives the desired functions $\tilde{Q}_{h:n}(\boldsymbol{\theta}_h)$.

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

$$Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}}) := Q_{\bar{m}:n}(\mathbf{Y}_{\bar{m}:1}, \dots, \mathbf{Y}_{\bar{m}:n}; \boldsymbol{\theta}_{\bar{m}}) := \sum_{h=1}^m \tilde{Q}_{h:n}(\boldsymbol{\theta}_h), \quad (\text{B.12a})$$

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

$$\boldsymbol{\theta}_{\bar{m}} := [\boldsymbol{\theta}'_1, \dots, \boldsymbol{\theta}'_m]'. \quad (\text{B.12b})$$

The m components $\tilde{Q}_{h:n}(\boldsymbol{\theta}_h)$ in the sum that defines $Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}})$ 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 $\{\mathbf{Y}_{\bar{m}:t}\}_{t \in \mathbb{Z}}$ and for a given bandwidth \mathbf{b} , the optimal parameter vector $\hat{\boldsymbol{\theta}}_{\bar{m}:n}$ for $Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}})$ 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 $\hat{\boldsymbol{\theta}}_{h:n}$ that shows up for the m bivariate cases in eq. (B.10). Since each $\hat{\boldsymbol{\theta}}_{h:n}$ converge almost surely to $\boldsymbol{\theta}_{h:\mathbf{b}}$, it is clear that $\hat{\boldsymbol{\theta}}_{\bar{m}:n}$ will converge almost surely to $\boldsymbol{\theta}_{\bar{m}:\mathbf{b}}$, the vector obtained by stacking the m vectors $\boldsymbol{\theta}_{h:\mathbf{b}}$ on top of each other.

The desired asymptotic result for the fixed \mathbf{b} and fixed m estimates $\hat{f}_v^m(\omega)$ can be obtained directly from the preceding observation and Theorem 1 in (Tjøstheim and Hufthammer, 2013), but that would not reveal how m and \mathbf{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 $Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}})$,

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

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

Remark B.2. The following matrix-observations gives the foundation for the extension from the bivariate case to the multivariate case.

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

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).

$$n^{-1}\nabla_{\bar{m}}Q_{\bar{m}:n}(\theta_{\bar{m}:b}) \xrightarrow{a.s.} \mathbf{0}$$

Proof. Since $\nabla_{\bar{m}}Q_{\bar{m}:n}(\theta_{\bar{m}:b})$ is the vector obtained by stacking the m vectors $\nabla_h Q_{h:n}(\theta_{h:b})$ on top of each other, and the proof of Theorem 1 in (Tjøstheim and Hufthammer, 2013) shows that $n^{-1}\nabla_h Q_{h:n}(\theta_{h:b})$ converges almost surely to $\mathbf{0}$, the same must necessarily be true for the combined vector $n^{-1}\nabla_{\bar{m}}Q_{\bar{m}:n}(\theta_{\bar{m}:b})$ too. \square

Lemma B.4 (Item (A2) of theorem B.1).

$$n^{-1}V_{\bar{m}|b:n} \xrightarrow{a.s.} V_{\bar{m}|b}, \text{ where } V_{\bar{m}|b} \text{ is a } 5m \times 5m \text{ positive definite matrix.}$$

Proof. Since $V_{\bar{m}|b:n}$ is the direct sum of the m matrices $V_{h:b:n}$, the behaviour of those will describe the behaviour of $V_{\bar{m}|b:n}$. The proof of Theorem 1 in (Tjøstheim and Hufthammer, 2013) shows that the matrices $n^{-1}V_{h:b:n}$ converges almost surely to positive definite matrices $V_{h:b}$, and this implies that $n^{-1}V_{\bar{m}|b:n}$ will converge almost surely to a block diagonal matrix $V_{\bar{m}|b}$, defined as the direct sum of the matrices $V_{h:b}$. 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, it follows that $V_{\bar{m}|b:n}$ is positive definite since all the $V_{h:b:n}$ are positive definite. \square

Lemma B.5 (Item (A3) of theorem B.1).

For $j, k = 1, \dots, 5m$,

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

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

Proof. $T_{\bar{m}|b:n}$ is the direct sum of the m matrices $T_{h:b:n}$, 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 $T_{h:b:n}$, which implies that it holds for $T_{\bar{m}|b:n}$ too. \square

Lemma B.6 (Item (A4) of theorem B.1.).

$$n^{-1/2} \nabla_{\bar{m}} Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}:b}) \xrightarrow{d} \mathbf{N}(\mathbf{0}, W_{\bar{m}|b})$$

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 $\nabla_{\bar{m}} Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}:b})$ by the help of 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 $\nabla_{\bar{m}} Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}:b})$ 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 $\nabla_{\bar{m}} Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}:b}) = [\nabla_1 Q_{1:n}(\boldsymbol{\theta}_{h:b})', \dots, \nabla_m Q_{m:n}(\boldsymbol{\theta}_{h:b})']'$, its components can be indexed by pairs $[h, i]$, $h = 1, \dots, m$ and $i = 1, \dots, 5$. From eq. (B.8) it is clear that the $[h, i]$ -component of the vector can be written as

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

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

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

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

The required α -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^ν -theory observed in eq. (C.41) gives $E[|X_{hi:t}|^\nu] < \infty$. Finally, since $\boldsymbol{\theta}_{h:b}$ is the population value parameter that minimise eq. (B.5), it follows that $E[X_{hi:t}] = 0$. 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 $S_{hi|n} := \sum_{t=1}^n X_{hi:t}$, Theorem 2.20(i) gives the asymptotic result

$$n^{-1} S_{hi|n} \longrightarrow \sigma^2 := \gamma_0 + 2 \sum_{\ell \geq 1} \gamma_\ell, \quad (\text{B.17})$$

with γ_ℓ being the ℓ^{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.

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

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

$$S_n(\mathbf{a}) = \sum_{t=1}^n X_t(\mathbf{a}), \quad (\text{B.19})$$

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

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

$$n^{-1} S_n(\mathbf{a}) \longrightarrow \sigma^2(\mathbf{a}) := \gamma_0(\mathbf{a}) + 2 \sum_{\ell \geq 1} \gamma_\ell(\mathbf{a}) \quad (\text{B.20})$$

$$n^{-1/2} S_n(\mathbf{a}) \xrightarrow{d} N(0, \sigma^2(\mathbf{a})), \quad (\text{B.21})$$

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

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

$$W_{\bar{m}|\mathbf{b}} := \text{Var}(\mathbf{X}_t) + 2 \sum_{\ell \geq 1} \text{Cov}(\mathbf{X}_{t+\ell}, \mathbf{X}_t) \quad (\text{B.22})$$

$$= E[\mathbf{X}_t \mathbf{X}_t'] + 2 \sum_{\ell \geq 1} E[\mathbf{X}_{t+\ell} \mathbf{X}_t'], \quad (\text{B.23})$$

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

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

Lemmas B.3 to B.6 shows that the penalty function $Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}})$ (for fixed m and fixed \mathbf{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

$$\sqrt{n} \left(\hat{\boldsymbol{\theta}}_{\bar{m}:n} - \boldsymbol{\theta}_{\bar{m}:\mathbf{b}} \right) \xrightarrow{d} N(\mathbf{0}, V_{\bar{m}|\mathbf{b}}^{-1} W_{\bar{m}|\mathbf{b}} V_{\bar{m}|\mathbf{b}}^{-1}). \quad (\text{B.24})$$

The hard task to deal with in the general situation, when $m \rightarrow \infty$ and $\mathbf{b} \rightarrow \mathbf{0}^+$ as $n \rightarrow \infty$, is the asymptotic behaviour of $n^{-1/2} \nabla_{\bar{m}} Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}:\mathbf{b}})$. 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 ‘ $m \rightarrow \infty$ and $\mathbf{b} \rightarrow \mathbf{0}^+$ when $n \rightarrow \infty$ ’, 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 $\mathbf{Y}_{h:t}$ from definition 2.6 will now be used to construct new random variables, that culminates in a random variable $\mathbf{Q}_{\bar{m}}^n$ which has the same limiting distribution²⁹ $\sqrt{b_1 b_2} \nabla_{\bar{m}} Q_{\bar{m};n}(\boldsymbol{\theta}_{\bar{m};b})$. Looking upon eq. (B.8), it is clear that everything depends upon the three functions $\psi(\mathbf{y}_h; \boldsymbol{\theta}_h)$, $\mathbf{u}_h(\mathbf{y}_h; \boldsymbol{\theta}_h)$ and $K_{h;b}(\mathbf{y}_h - \mathbf{v})$. The number of parameters in $\boldsymbol{\theta}_h$ will henceforth be denoted with p , since the discussion needs to encompass both $p = 1$ and $p = 5$.

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

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

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

(b) For $L \geq 0$, define the following lower and upper truncated versions of $u_{h;q;b}(\mathbf{w})$,

$$u_{h;q;b}(\mathbf{w})^{\leq L} := u_{h;q;b}(\mathbf{w}) \cdot \mathbb{1}\{|u_{h;q;b}(\mathbf{w})| \leq L\}, \quad (\text{B.26a})$$

$$u_{h;q;b}(\mathbf{w})^{>L} := u_{h;q;b}(\mathbf{w}) \cdot \mathbb{1}\{|u_{h;q;b}(\mathbf{w})| > L\}. \quad (\text{B.26b})$$

Obviously; $u_{h;q;b}(\mathbf{w}) = u_{h;q;b}(\mathbf{w})^{\leq L} + u_{h;q;b}(\mathbf{w})^{>L}$ and $u_{h;q;b}(\mathbf{w})^{\leq L} \cdot u_{h;q;b}(\mathbf{w})^{>L} = 0$.

(c) Let $u_{h;q}(\mathbf{w})$ be as in item (a), with the difference that the limit $\mathbf{b} \rightarrow \mathbf{0}^+$ of the parameters $\boldsymbol{\theta}_{h;b}$ are used in the definition.³⁰ Let $u_{h;q}(\mathbf{w})^{\leq L}$ and $u_{h;q}(\mathbf{w})^{>L}$ be the truncated versions of $u_{h;q}(\mathbf{w})$.

The following simple observations will be useful later on.

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

(a) $\sup_{h;q} |u_{h;q;b}(\mathbf{v})| < \infty$ and $\sup_{h;q} |u_{h;q}(\mathbf{v})| < \infty$.

(b) When L is large enough, $u_{h;q;b}(\mathbf{v})^{\leq L} = u_{h;q;b}(\mathbf{v})$ and $u_{h;q}(\mathbf{v})^{\leq L} = u_{h;q}(\mathbf{v})$.

Proof. By definition, the functions $u_{h;q;b}(\mathbf{w})$ and $u_{h;q}(\mathbf{w})$ 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 \mathbf{v} . Since the parameters in these polynomials originates from a local Gaussian approximation of $g_h(\mathbf{y}_h)$ at the point \mathbf{v} , and since assumption 2.1(b) ensures that the bivariate densities $g_h(\mathbf{y}_h)$ will approach the product of the marginal densities when $h \rightarrow \infty$, 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). \square

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

²⁹Due to the presence of the kernel function $K_{h;b}(\mathbf{w})$, the fourth requirement of the Klimko-Nelson approach will (when $\mathbf{b} \rightarrow \mathbf{0}^+$) 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.

³⁰The limit of the parameters $\boldsymbol{\theta}_{h;b}$ will exist under assumptions that implies that the four requirements of the Klimko-Nelson approach are satisfied, cf. Tjøstheim and Hufthammer (2013) for details.

normal kernels. The following definition enables a more general approach to be used in the theoretical investigation,³¹ 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 $K(\mathbf{w})$, that satisfies*

$$\int_{\mathbb{R}^2} K(w_1, w_2) \, dw_1 \, dw_2 = 1, \quad (\text{B.27a})$$

$$\mathcal{K}_{1:k}(w_2) := \int_{\mathbb{R}^1} K(w_1, w_2) w_1^k \, dw_1 \quad \text{is bounded for } k \in \{0, 1, 2\}, \quad (\text{B.27b})$$

$$\mathcal{K}_{2:\ell}(w_1) := \int_{\mathbb{R}^1} K(w_1, w_2) w_2^\ell \, dw_2 \quad \text{is bounded for } \ell \in \{0, 1, 2\}, \quad (\text{B.27c})$$

$$\int_{\mathbb{R}^2} K(w_1, w_2) |w_1^k w_2^\ell| \, dw_1 \, dw_2 < \infty, \quad k, \ell \geq 0 \text{ and } k + \ell \leq 2 \cdot \lceil \nu \rceil, \quad (\text{B.27d})$$

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

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

It turns out, see appendix C.4 for details, that the asymptotic results needed later on mainly depends upon the properties of the kernel $K(\mathbf{w})$ and the components $u_{hq:b}(\mathbf{w})$ 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 $g_h(\mathbf{y}_h)$, $u_{hq:b}(\mathbf{w})$ and $K(\mathbf{w})$ as given in definitions 2.6, B.7 and B.9, let $\mathfrak{U}_{h:b} := [u_{h1:b}(\mathbf{v}), \dots, u_{hp:b}(\mathbf{v})]'$, and define the following matrices.*

$$W_{h:b} := \mathfrak{U}_{h:b} \mathfrak{U}'_{h:b} \cdot g_h(\mathbf{v}) \int_{\mathbb{R}^2} K(\mathbf{w})^2 \, d\mathbf{w}, \quad (\text{B.29a})$$

$$W_{\bar{m}|b} := \bigoplus_{h=1}^m W_{h:b}. \quad (\text{B.29b})$$

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

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

Definition B.11. *Based on $\mathbf{Y}_{h:t}$, $u_{hq:b}(\mathbf{w})$ and $K_{h:b}(\mathbf{y}_h - \mathbf{v})$ from definitions 2.6, B.7 and B.9, define new bivariate random variables as follows,*

$$X_{hq:t}^n(\mathbf{v}) := \sqrt{b_1 b_2} K_{h:b}(\mathbf{Y}_{h:t} - \mathbf{v}) u_{hq:b}(\mathbf{Y}_{h:t}), \quad (\text{B.30a})$$

$$X_{hq:t}^{n|\leq L}(\mathbf{v}) := \sqrt{b_1 b_2} K_{h:b}(\mathbf{Y}_{h:t} - \mathbf{v}) u_{hq:b}(\mathbf{Y}_{h:t})^{\leq L}, \quad (\text{B.30b})$$

$$X_{hq:t}^{n|>L}(\mathbf{v}) := \sqrt{b_1 b_2} K_{h:b}(\mathbf{Y}_{h:t} - \mathbf{v}) u_{hq:b}(\mathbf{Y}_{h:t})^{>L}. \quad (\text{B.30c})$$

³¹Differences in the computational cost implies that the product normal kernel is used for practical purposes.

Obviously; $X_{hq:t}^n(\mathbf{v}) = X_{hq:t}^{n|\leq L}(\mathbf{v}) + X_{hq:t}^{n|>L}(\mathbf{v})$ and $X_{hq:t}^{n|\leq L}(\mathbf{v}) \cdot X_{hq:t}^{n|>L}(\mathbf{v}) = 0$.

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

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

$$\tilde{X}_{hq:t}^n := X_{hq:t}^n - \sqrt{b_1 b_2} \int_{\mathbb{R}^2} K_{h:b}(\mathbf{y}_h - \mathbf{v}) u_{hq:b}(\mathbf{y}_h) \psi(\mathbf{y}_h; \boldsymbol{\theta}_h) d\mathbf{y}_h, \quad (\text{B.31})$$

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 $Z_{hq:t}^n$ and \mathcal{Q}_{hq}^n below, and the only place $\tilde{X}_{hq:t}^n$ is needed is in the proof of lemma B.14.

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

$$Z_{hq:t}^n := X_{hq:t}^n - \mathbb{E}[X_{hq:t}^n], \quad (\text{B.32a})$$

$$\mathcal{Q}_{hq}^n := \sum_{t=1}^n Z_{hq:t}^n. \quad (\text{B.32b})$$

Similarly, $Z_{hq:t}^{n|\geq L}$, $Z_{hq:t}^{n|<L}$, $\mathcal{Q}_{hq}^{n|\geq L}$ and $\mathcal{Q}_{hq}^{n|<L}$ can be defined in the natural manner, with the obvious connections $Z_{hq:t}^n = Z_{hq:t}^{n|\geq L} + Z_{hq:t}^{n|<L}$, $Z_{hq:t}^{n|\geq L} \cdot Z_{hq:t}^{n|<L} = 0$, and $\mathcal{Q}_{hq}^n = \mathcal{Q}_{hq}^{n|\geq L} + \mathcal{Q}_{hq}^{n|<L}$ holding for all L . Moreover: $\text{Cov}(Z_{hq:i}^n, Z_{j:k}^n) = \mathbb{E}[Z_{hq:i}^n \cdot Z_{j:k}^n] = \text{Cov}(X_{hq:i}^n, X_{j:r:k}^n)$.

The last batch of random variables can now be introduced.

Definition B.13. Based upon the bivariate $Z_{hq:t}^n$ from definition B.12, and for $\mathbf{a} := \mathbf{a}_{\bar{m}} \in \mathbb{R}^{p \times m}$, define the following $(m+1)$ -variate random variables,

$$Z_{\bar{m}:t}^n(\mathbf{a}) := \sum_{h=1}^m \sum_{q=1}^p a_{hq} Z_{hq:t}^n = \mathbf{a}' \mathbf{Z}_{\bar{m}:t}^n, \quad (\text{B.33a})$$

$$\mathcal{Q}_{\bar{m}}^n(\mathbf{a}) := \sum_{h=1}^m \sum_{q=1}^p a_{hq} \mathcal{Q}_{hq}^n = \mathbf{a}' \mathcal{Q}_{\bar{m}}^n, \quad (\text{B.33b})$$

where $\mathbf{Z}_{\bar{m}:t}^n$ and $\mathcal{Q}_{\bar{m}}^n$ are defined in the obvious manner.

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

Proof. The only difference between $\mathcal{Q}_{\bar{m}}^n$ and $\sqrt{b_1 b_2} \nabla_{\bar{m}} Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}:b})$ is that the first use $Z_{hq:t}^n$ where the second use $\tilde{X}_{hq:t}^n$. The difference between these components are

$$Z_{hq:t}^n - \tilde{X}_{hq:t}^n = \sqrt{b_1 b_2} \cdot \int_{\mathbb{R}^2} K_{h:b}(\mathbf{y}_h - \mathbf{v}) u_{hq:b}(\mathbf{y}_h) \{g_h(\mathbf{y}_h) - \psi(\mathbf{y}_h; \boldsymbol{\theta}_h)\} d\mathbf{y}_h, \quad (\text{B.34})$$

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

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 $m \rightarrow \infty$ and $\mathbf{b} \rightarrow \mathbf{0}^+$ when $n \rightarrow \infty$.

Compared to the discussion in appendices B.1.2 and B.1.3, the effect of p free parameters instead of 5 free parameters is that the $m \times m$ block-matrices will have components that are matrices of size $p \times p$ instead of size 5×5 – except for this, the arguments in lemmas B.3 to B.5 will be unaffected, i.e. the three first assumptions of the Klimko-Nelson approach have already been covered.

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 $n^{-1/2} \mathbf{Q}_{\bar{m}}^n$ approaches a normal distribution when \mathbf{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 $\mathbb{E}[X_{hq:i}^n \cdot X_{jr:k}^n]$. The computation of this expectation will (depending on the indices h, i, j and k) either require a bivariate, trivariate or tetravariate integral.

Combinations	\mathbf{v}	\mathbf{b}	$\mathbf{Y}_{h:i}$	$\mathbf{Y}_{j:k}$
First argument of $K_{h:b}$	v_1	b_1	Y_{h+i}	Y_{j+k}
Second argument of $K_{h:b}$	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 $\mathbb{E}[X_{hq:i}^n \cdot X_{jr:k}^n]$, i.e. the presence of \mathbf{v} and \mathbf{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. (2.13) of definition 2.4(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 $K(\mathbf{w})$ or the score-function components $u_{h_q:b}(\mathbf{w})$ 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, $K(\mathbf{w})$ and $u_{h_q:b}(\mathbf{w})$ 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 $u_{h_q:b}(\mathbf{w})$ and $K(\mathbf{w})$ are as given in definitions B.7 and B.9, then the random variables $X_{hq:t}^n$ from definition B.11 satisfies*

- (a) $\mathbb{E}[X_{hq:i}^n] = O(\sqrt{b_1 b_2})$.
- (b) $\mathbb{E}[|X_{hq:i}^n|^\nu]^{1/\nu} = O(|b_1 b_2|^{(2-\nu)/2\nu})$.

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

where bivariate, trivariate and tetrivariate 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 $u_{hq:b}(\mathbf{w})$ and $K(\mathbf{w})$. 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 tetrivariate cases of item (c).

Equation (2.18) from assumption 2.1(d) is needed for the bivariate case of item (c), i.e. the Taylor expansion of $g_h(\mathbf{y}_h)$ around the point \mathbf{v} allows the integral of interest to be written as the sum of the following three integrals:

$$\mathcal{J}_1 := \int_{\mathbb{R}^2} \mathcal{V}(\mathbf{y}_h) \cdot g_h(\mathbf{v}) d\mathbf{y}_h, \quad (\text{B.35a})$$

$$\mathcal{J}_2 := \int_{\mathbb{R}^2} \mathcal{V}(\mathbf{y}_h) \cdot (\mathbf{g}_h(\mathbf{v})' [\mathbf{y}_h - \mathbf{v}]) d\mathbf{y}_h, \quad (\text{B.35b})$$

$$\mathcal{J}_3 := \int_{\mathbb{R}^2} \mathcal{V}(\mathbf{y}_h) \cdot (\mathfrak{R}_h(\mathbf{y}_h)' [\mathbf{y}_h - \mathbf{v}]) d\mathbf{y}_h. \quad (\text{B.35c})$$

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 $w_2 = (y_h - v_2)/b_2$ must be applied, which in particular replaces the vector $[\mathbf{y}_h - \mathbf{v}]$ 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 $\mathbf{g}_h(\mathbf{v})$, let \mathcal{W} be the substituted version of \mathcal{V} , let \mathfrak{R}_{h_1} and \mathfrak{R}_{h_2} be the two components of the remainder function and finally let \mathfrak{T}_{h_1} and \mathfrak{T}_{h_2} be the substituted versions of $\mathfrak{R}_{h_1} \mathcal{W}$ and $\mathfrak{R}_{h_2} \mathcal{W}$.

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

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

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.

$$\mathcal{J}_3 = \int_{\mathbb{R}^2} [b_1 w_1 \cdot \mathfrak{T}_{h_1}(\mathbf{y}(\mathbf{w})) + b_2 w_2 \cdot \mathfrak{T}_{h_2}(\mathbf{y}(\mathbf{w}))] d\mathbf{w}, \quad (\text{B.37})$$

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 \square

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

Corollary B.16. *When Y_t satisfies assumption 2.1, and $u_{h_q;b}(\mathbf{w})$ and $K(\mathbf{w})$ are as given in definitions B.7 and B.9, then the random variables $X_{h_q;t}^n$ from definition B.11 satisfies*

$$\text{Cov}(X_{h_q;i}^n, X_{j_r;k}^n) = \begin{cases} u_{h_q;b}(\mathbf{v}) u_{h_r;b}(\mathbf{v}) g_h(\mathbf{v}) \int_{\mathbb{R}^2} K(\mathbf{w})^2 d\mathbf{w} + O(b_1 \vee b_2) & \text{when bivariate,} \\ O(b_1 \wedge b_2) & \text{when trivariate,} \\ O(b_1 b_2) & \text{when tetrivariate.} \end{cases} \quad (\text{B.38})$$

Proof. Since $\text{Cov}(X_{h_q;i}^n, X_{j_r;k}^n) = \mathbb{E}[X_{h_q;i}^n \cdot X_{j_r;k}^n] - \mathbb{E}[X_{h_q;i}^n] \cdot \mathbb{E}[X_{j_r;k}^n]$, the result follows immediately from an inspection of items (a) and (c) of lemma B.15. \square

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

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

- (a) $\mathbb{E}[|Z_{h_q;t}^n|^\nu]^{1/\nu} = O(|b_1 b_2|^{(2-\nu)/2\nu})$.
- (b) $\mathbb{E}[|Z_{\bar{m};t}^n(\mathbf{a})|^\nu]^{1/\nu} = O(m |b_1 b_2|^{(2-\nu)/2\nu})$.

Proof. The connection between expectations and L^ν -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

$$\mathbb{E}[|Z_{h_q;t}^n|^\nu]^{1/\nu} = \mathbb{E}[|X_{h_q;t}^n - \mathbb{E}[X_{h_q;t}^n]|^\nu]^{1/\nu} \quad (\text{B.39a})$$

$$\leq \mathbb{E}[|X_{h_q;t}^n|^\nu]^{1/\nu} + \mathbb{E}[|\mathbb{E}[X_{h_q;t}^n]|^\nu]^{1/\nu} \quad (\text{B.39b})$$

$$= O(|b_1 b_2|^{(2-\nu)/2\nu}) + O(\sqrt{|b_1 b_2|}) \quad (\text{B.39c})$$

$$= O(|b_1 b_2|^{(2-\nu)/2\nu}). \quad (\text{B.39d})$$

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

$$\mathbb{E}[|Z_{\bar{m};t}^n(\mathbf{a})|^\nu]^{1/\nu} = \mathbb{E}\left[\left|\sum_{h=1}^m \sum_{q=1}^p a_{h_q} Z_{h_q;t}^n\right|^\nu\right]^{1/\nu} \quad (\text{B.40a})$$

$$\leq \sum_{h=1}^m \sum_{q=1}^p |a_{h_q}| \mathbb{E}[|Z_{h_q;t}^n|^\nu]^{1/\nu} \quad (\text{B.40b})$$

$$\leq \sum_{h=1}^m \sum_{q=1}^p A_{\bar{m}} \cdot O(|b_1 b_2|^{(2-\nu)/2\nu}) \quad (\text{B.40c})$$

$$= O(m |b_1 b_2|^{(2-\nu)/2\nu}). \quad (\text{B.40d})$$

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

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 $m \rightarrow \infty$ and $\mathbf{b} \rightarrow \mathbf{0}^+$ when $n \rightarrow \infty$. 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 \mathbf{b} are as specified in assumption 2.3, and when $u_{h,q;\mathbf{b}}(\mathbf{w})$ and $K(\mathbf{w})$ are as given in definitions B.7 and B.9 – then the random variables $Z_{\bar{m};t}^n(\mathbf{a})$ from definition B.13 satisfies*

$$\frac{1}{n} \sum_{\substack{i,k=1 \\ i \neq k}}^n |\mathbb{E}[Z_{\bar{m};i}^n(\mathbf{a}) \cdot Z_{\bar{m};k}^n(\mathbf{a})]| = o(1). \quad (\text{B.41})$$

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.

$$\frac{1}{n} \sum_{\substack{i,k=1 \\ i \neq k}}^n |\mathbb{E}[Z_{\bar{m};i}^n(\mathbf{a}) \cdot Z_{\bar{m};k}^n(\mathbf{a})]| = 2 \sum_{\ell=1}^{n-1} \left(1 - \frac{\ell}{n}\right) I_{\bar{m};\ell}^n(\mathbf{a}), \quad (\text{B.42})$$

where the terms $I_{\bar{m};\ell}^n(\mathbf{a})$ are given by

$$I_{\bar{m};\ell}^n(\mathbf{a}) := |\mathbb{E}[Z_{\bar{m};0}^n(\mathbf{a}) \cdot Z_{\bar{m};\ell}^n(\mathbf{a})]| \quad (\text{B.43a})$$

$$= \left| \mathbb{E} \left[\sum_{h=1}^m \sum_{q=1}^p a_{hq} Z_{hq;0}^n \cdot \sum_{j=1}^m \sum_{r=1}^p a_{jr} Z_{jr;\ell}^n \right] \right| \quad (\text{B.43b})$$

$$= \left| \sum_{h=1}^m \sum_{j=1}^m \sum_{q=1}^p \sum_{r=1}^p a_{hq} a_{jr} \mathbb{E}[Z_{hq;0}^n \cdot Z_{jr;\ell}^n] \right| \quad (\text{B.43c})$$

$$\leq \sum_{h=1}^m \sum_{j=1}^m \sum_{q=1}^p \sum_{r=1}^p |a_{hq}| |a_{jr}| I_{hqjr;\ell}^n, \quad (\text{B.43d})$$

where $I_{hqjr;\ell}^n := |\mathbb{E}[Z_{hq;0}^n \cdot Z_{jr;\ell}^n]| = |\text{Cov}(X_{hq;0}^n, X_{jr;\ell}^n)|$.

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

$$J_1 := 2 \sum_{\ell=1}^m (1 - \ell/n) I_{\bar{m}:\ell}^n(\mathbf{a}), \quad (\text{B.44a})$$

$$J_2 := 2 \sum_{\ell=m+1}^{k_n+m} (1 - \ell/n) I_{\bar{m}:\ell}^n(\mathbf{a}), \quad (\text{B.44b})$$

$$J_3 := 2 \sum_{\ell=k_n+m+1}^{n-1} (1 - \ell/n) I_{\bar{m}:\ell}^n(\mathbf{a}). \quad (\text{B.44c})$$

From the definition of $I_{\bar{m}:\ell}^n(\mathbf{a})$ it is seen that in J_1 there will be some overlap between those Y_t that are a part of $Z_{\bar{m}:0}^n(\mathbf{a})$ and those that are a part of $Z_{\bar{m}:\ell}^n(\mathbf{a})$, 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.

$$0 \leq J_1 \leq 2 \cdot \left(\max_{\substack{h \in \{1, \dots, m\} \\ q \in \{1, \dots, p\}}} |a_{hq}|^2 \right) \cdot \sum_{\ell=1}^m \sum_{h=1}^m \sum_{j=1}^m \sum_{q=1}^p \sum_{r=1}^p |\text{Cov}(X_{hq;0}^n, X_{jr;\ell}^n)|, \quad (\text{B.45})$$

and corollary B.16 can be used to determine how the summand behaves in the limit. Table 1, page 53, shows that the bivariate case never occurs, that h must be equal to ℓ 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;

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

$$m^3 b_1 b_2 \leq m^{-1} \cdot m^4(b_1 \vee b_2)^2 = m^{-1} \cdot (m^2(b_1 \vee b_2))^2. \quad (\text{B.46b})$$

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

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

Since $k_n m^2 b_1 b_2 \rightarrow 0$ (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 $I_{\bar{m}:\ell}^n(\mathbf{a})$, 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 \mathbf{b}) the requirement about finite expectations follows from corollary B.17(b).

The σ -algebras to be used follows from the comment stated after eq. (C.33), i.e. that $Z_{\bar{m}:0}^n(\mathbf{a}) \in \mathcal{F}_0^m$, whereas $Z_{\bar{m}:\ell}^n(\mathbf{a}) \in \mathcal{F}_\ell^{\ell+m} \subset \mathcal{F}_{m+(\ell-m)}^\infty$. Thus, for $\ell > k_n + m$, the following bound is obtained on $I_{\bar{m}:\ell}^n(\mathbf{a})$,

$$I_{\bar{m}:\ell}^n(\mathbf{a}) = |\mathbb{E}[Z_{\bar{m}:0}^n(\mathbf{a}) \cdot Z_{\bar{m}:\ell}^n(\mathbf{a})]| \quad (\text{B.48a})$$

$$= |\mathbb{E}[Z_{\bar{m}:0}^n(\mathbf{a}) \cdot Z_{\bar{m}:\ell}^n(\mathbf{a})] - \mathbb{E}[Z_{\bar{m}:0}^n(\mathbf{a})] \cdot \mathbb{E}[Z_{\bar{m}:\ell}^n(\mathbf{a})]| \quad (\text{B.48b})$$

$$\leq 12 (\mathbb{E}[|Z_{\bar{m}:0}^n(\mathbf{a})|^\nu])^{1/\nu} \cdot (\mathbb{E}[|Z_{\bar{m}:\ell}^n(\mathbf{a})|^\nu])^{1/\nu} \cdot [\alpha(\ell - m)]^{1-1/\nu-1/\nu} \quad (\text{B.48c})$$

$$= 12 ((\mathbb{E}[|Z_{\bar{m}:0}^n(\mathbf{a})|^\nu])^{1/\nu})^2 \cdot [\alpha(\ell - m)]^{1-2/\nu} \quad (\text{B.48d})$$

$$= 12 (O(m |b_1 b_2|^{(2-\nu)/2\nu}))^2 \cdot [\alpha(\ell - m)]^{1-2/\nu} \quad (\text{B.48e})$$

$$\leq \mathcal{C} \cdot m^2 \cdot |b_1 b_2|^{(2-\nu)/\nu} \cdot [\alpha(\ell - m)]^{1-2/\nu}, \quad (\text{B.48f})$$

where eq. (B.48b) follows since the mean of $Z_{\bar{m}:t}^n(\mathbf{a})$ 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

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

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^{th} root on both sides, it is clear that the desired inequality can be obtained when $k_n + 1 = \lceil m^{2/a} \cdot |b_1 b_2|^{(2-\nu)/a\nu} \rceil$, which gives the new bound

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

and if $k_n \rightarrow \infty$ when $n \rightarrow \infty$, 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 $k_n m^2 b_1 b_2 \rightarrow 0$ (needed for the J_2 -term) and $k_n \rightarrow \infty$ (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. \square

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 \mathbf{b} are as specified in assumption 2.3, and with $W_{\bar{m}:\mathbf{b}} = \bigoplus_{h=1}^m W_{h:\mathbf{b}}$ and $\mathbf{a} = \mathbf{a}_{\bar{m}} = [\mathbf{a}_1, \dots, \mathbf{a}_m]'$ (with $\mathbf{a}_h \in \mathbb{R}^p$) as given in definition B.10, then the random variable $Z_{\bar{m}:t}^n(\mathbf{a})$ from definition B.13 satisfies*

$$(a) \text{Var}(Z_{\bar{m}:t}^n(\mathbf{a})) = \mathbf{a}'_{\bar{m}} W_{\bar{m}:\mathbf{b}} \mathbf{a}_{\bar{m}} + O(m^2 \cdot (b_1 \vee b_2)) = \sum_{h=1}^m \mathbf{a}'_h W_{h:\mathbf{b}} \mathbf{a}_h + O(m^2 \cdot (b_1 \vee b_2)) = O(m).$$

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

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

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

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

$$\text{Var}(Z_{\bar{m}:t}^n(\mathbf{a})) = \sum_{h=1}^m \sum_{j=1}^m \sum_{q=1}^p \sum_{r=1}^p a_{hq} a_{jr} \text{Cov}(X_{hq:t}^n, X_{jr:t}^n) \quad (\text{B.51a})$$

$$= \sum_{h=1}^m \sum_{q=1}^p \sum_{r=1}^p a_{hq} a_{hr} \text{Cov}(X_{hq:t}^n, X_{hr:t}^n) + \sum_{\substack{h,j=1 \\ h \neq j}}^m \sum_{q=1}^p \sum_{r=1}^p a_{hq} a_{jr} \text{Cov}(X_{hq:t}^n, X_{jr:t}^n). \quad (\text{B.51b})$$

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 tetrivariate cases can be applied to the ‘off-diagonal part’. The ‘diagonal part’ can thus be written as the sum of $\sum_{h=1}^m \sum_{q=1}^p \sum_{r=1}^p a_{hq} a_{hr} u_{hq;b}(\mathbf{v}) u_{hr;b}(\mathbf{v}) g_h(\mathbf{v}) \int_{\mathbb{R}^2} K(\mathbf{w})^2 d\mathbf{w}$ (which is equal to $\mathbf{a}' W_{\bar{m}:b} \mathbf{a} = \sum_{h=1}^m \mathbf{a}'_h W_{h:b} \mathbf{a}_h$) 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 $\mathbf{a}'_h W_{h:b} \mathbf{a}_h$ are finite.

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

$$\text{Var}(\eta_{1:r}) = \sum_{i=1}^r \text{Var}(Z_{\bar{m}:i}^n(\mathbf{a})) + \sum_{\substack{i,k=1 \\ i \neq k}}^r \text{E}[Z_{\bar{m}:i}^n(\mathbf{a}) \cdot Z_{\bar{m}:k}^n(\mathbf{a})]. \quad (\text{B.52})$$

The ‘on diagonal’ part of this sum equals $r \cdot \text{Var}(Z_{\bar{m}:1}^n(\mathbf{a}))$ 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 $u_{hq;b}(\mathbf{v}) u_{hr;b}(\mathbf{v})$ that occurs in $W_{h:b}$ either are replaced by $u_{hq;b}(\mathbf{v})^{\leq L} u_{hr;b}(\mathbf{v})^{\leq L}$ or by $u_{hq;b}(\mathbf{v})^{>L} u_{hr;b}(\mathbf{v})^{>L}$. Lemma B.8(b) gives that $u_{hq;b}(\mathbf{v})^{\leq L} = u_{hq;b}(\mathbf{v})$ when L is large enough (and thus $u_{hq;b}(\mathbf{v})^{>L} = 0$), which completes the proof. \square

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 $Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}:b})$, from which it follows an asymptotic normality result for $\hat{\boldsymbol{\theta}}_{v|\bar{m}|b|p}$, that finally gives the asymptotic normality result of $\hat{f}_{v|p}^m(\omega)$. (Confer remark B.3 for an interpretation of the m that occurs in the limiting distributions.)

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

- (a) $n^{-1/2} \mathcal{Q}_{\bar{m}}^n(\mathbf{a}) \xrightarrow{d} \mathbf{N}(0, \sum_{h=1}^m \mathbf{a}'_h W_{h:b} \mathbf{a}_h)$, i.e. asymptotically univariate normal.
 (b) $n^{-1/2} \mathcal{Q}_{\bar{m}}^n \xrightarrow{d} \mathbf{N}(\mathbf{0}, \bigoplus_{h=1}^m W_{h:b})$, i.e. asymptotically mp-variate normal.

Proof. For the proof of item (a), note the following connection between $\mathcal{Q}_{\bar{m}}^n(\mathbf{a})$ and $Z_{\bar{m}:t}^n(\mathbf{a})$ which follows directly from definitions B.12 and B.13,

$$\mathcal{Q}_{\bar{m}}^n(\mathbf{a}) = \sum_{h=1}^m \sum_{q=1}^p a_{hq} \mathcal{Q}_{hq}^n = \sum_{h=1}^m \sum_{q=1}^p a_{hq} \left[\sum_{t=1}^n Z_{hq:t}^n \right] = \sum_{t=1}^n \left[\sum_{h=1}^m \sum_{q=1}^p a_{hq} Z_{hq:t}^n \right] = \sum_{t=1}^n Z_{\bar{m}:t}^n(\mathbf{a}). \quad (\text{B.53a})$$

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

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

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

$$\mathcal{B}_j := \{(j-1)(r+s) + r + 1, \dots, j(r+s)\}, \text{ for } j = 1, \dots, \ell, \quad (\text{B.54b})$$

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

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

$$\mathcal{A}^\circ := \bigcup_{j=1}^{\ell} \mathcal{A}_j, \quad \mathcal{B}^\circ := \bigcup_{j=1}^{\ell} \mathcal{B}_j. \quad (\text{B.55a})$$

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

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

$$\eta_j := \sum_{t \in \mathcal{A}_j} Z_{\bar{m}:t}^n(\mathbf{a}), \text{ for } j = 1, \dots, \ell, \quad S_n^{(1)} := \sum_{j=1}^{\ell} \eta_j = \sum_{t \in \mathcal{A}^\circ} Z_{\bar{m}:t}^n(\mathbf{a}), \quad (\text{B.56a})$$

$$\xi_j := \sum_{t \in \mathcal{B}_j} Z_{\bar{m}:t}^n(\mathbf{a}), \text{ for } j = 1, \dots, \ell, \quad S_n^{(2)} := \sum_{j=1}^{\ell} \xi_j = \sum_{t \in \mathcal{B}^\circ} Z_{\bar{m}:t}^n(\mathbf{a}), \quad (\text{B.56b})$$

$$\zeta_\ell := \sum_{t \in \mathcal{C}_\ell} Z_{\bar{m}:t}^n(\mathbf{a}), \quad S_n^{(3)} := \zeta_\ell, \quad (\text{B.56c})$$

such that

$$n^{-1/2} \mathcal{Q}_{\bar{m}}^n(\mathbf{a}) = n^{-1/2} \{S_n^{(1)} + S_n^{(2)} + S_n^{(3)}\}. \quad (\text{B.57})$$

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

$$\text{Var}(n^{-1/2} \mathcal{Q}_m^n(\mathbf{a})) = \frac{1}{n} \mathbb{E}[\mathcal{Q}_m^n(\mathbf{a}) \cdot \mathcal{Q}_m^n(\mathbf{a})] = \frac{1}{n} \sum_{p=1}^3 \sum_{q=1}^3 \mathbb{E}[S_n^{(p)} \cdot S_n^{(q)}]. \quad (\text{B.58})$$

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

$$\left| \frac{1}{n} \mathbb{E}[S_n^{(1)} \cdot S_n^{(2)}] \right| = \left| \frac{1}{n} \mathbb{E} \left[\left(\sum_{i \in \mathcal{A}^\circ} Z_{\bar{m}:i}^n(\mathbf{a}) \right) \cdot \left(\sum_{k \in \mathcal{B}^\circ} Z_{\bar{m}:k}^n(\mathbf{a}) \right) \right] \right| \quad (\text{B.59a})$$

$$\leq \frac{1}{n} \sum_{i \in \mathcal{A}^\circ} \sum_{k \in \mathcal{B}^\circ} |\mathbb{E}[Z_{\bar{m}:i}^n(\mathbf{a}) \cdot Z_{\bar{m}:k}^n(\mathbf{a})]| \quad (\text{B.59b})$$

$$\leq \frac{1}{n} \sum_{\substack{i,k=1 \\ i \neq k}}^n |\mathbb{E}[Z_{\bar{m}:i}^n(\mathbf{a}) \cdot Z_{\bar{m}:k}^n(\mathbf{a})]|. \quad (\text{B.59c})$$

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

$$\frac{1}{n} \mathbb{E}[S_n^{(2)} \cdot S_n^{(2)}] = \frac{1}{n} \sum_{i,k \in \mathcal{B}^\circ} \mathbb{E}[Z_{\bar{m}:i}^n(\mathbf{a}) \cdot Z_{\bar{m}:k}^n(\mathbf{a})] \quad (\text{B.60a})$$

$$= \frac{1}{n} \sum_{i \in \mathcal{B}^\circ} \mathbb{E}[Z_{\bar{m}:i}^n(\mathbf{a}) \cdot Z_{\bar{m}:i}^n(\mathbf{a})] + \frac{1}{n} \sum_{\substack{i,k \in \mathcal{B}^\circ \\ i \neq k}} \mathbb{E}[Z_{\bar{m}:i}^n(\mathbf{a}) \cdot Z_{\bar{m}:k}^n(\mathbf{a})] \quad (\text{B.60b})$$

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

$$= O\left(\frac{m\ell s}{n}\right). \quad (\text{B.60d})$$

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

$$\frac{1}{n} \mathbb{E}[S_n^{(3)} \cdot S_n^{(3)}] = O\left(\frac{m(n - \ell(r + s))}{n}\right) < O\left(\frac{mr}{n}\right). \quad (\text{B.61})$$

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 $n^{-1/2}(\mathcal{Q}_m^n(\mathbf{a}) - S_n^{(1)}) \Rightarrow 0$, and (Billingsley, 2012, Theorem 25.4) states that there thus is a common limiting distribution for $n^{-1/2} \mathcal{Q}_m^n(\mathbf{a})$ 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,

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

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

$$\left| \mathbb{E}[\exp(itS_n^{(1)})] - \prod_{j=1}^{\ell} \mathbb{E}[\exp(it\eta_j)] \right| \rightarrow 0. \quad (\text{B.63})$$

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

$$\vartheta := \{(j+1)(r+s)+1\} - \{(j-1)(r+s)+r+m\} = s - m + 1. \quad (\text{B.64})$$

Assumption 2.3(f), i.e. $m = o(s)$, ensures that there (asymptotically) will be no overlap between these σ -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 η_j are asymptotically independent.

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

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

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

$$\forall \epsilon > 0 \quad \lim_{n \rightarrow \infty} \sum_{j=1}^{\ell} \frac{1}{\mathfrak{s}_\ell^2} \mathbb{E}[\eta_j^2 \cdot \mathbb{1}\{|\eta_j| \geq \epsilon \sqrt{\mathfrak{s}_\ell^2}\}] \rightarrow 0. \quad (\text{B.66})$$

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

Keeping in mind the definitions of $X_{hq;t}^n$, $Z_{hq;t}^n$ and η_j , see eqs. (B.30a), (B.32a) and (B.56a), it is clear that an upper bound for $|\eta_j|$ might be deduced from,

$$|\eta_j| = \left| \sum_{t \in \mathcal{A}_j} \sum_{h=1}^m \sum_{q=1}^p a_{hq} Z_{hq;t}^n \right| \leq \sum_{t \in \mathcal{A}_j} \sum_{h=1}^m \sum_{q=1}^p |a_{hq}| |Z_{hq;t}^n|, \quad (\text{B.67a})$$

$$|Z_{hq;t}^n| = |X_{hq;t}^n - \mathbb{E}[X_{hq;t}^n]| \leq |X_{hq;t}^n| + O\left(\sqrt{b_1 b_2}\right), \quad (\text{B.67b})$$

$$|X_{hq;t}^n| = \left| \sqrt{b_1 b_2} \cdot \frac{1}{b_1 b_2} K_h \left(\frac{Y_{t+h} - v_1}{b_1}, \frac{Y_t - v_2}{b_2} \right) u_{h;b}(\mathbf{Y}_{h;t}) \right|. \quad (\text{B.67c})$$

If all of the functions $u_{h_q;b}(\mathbf{w})$ are bounded, or if the kernel functions $K_{h;b}(\mathbf{w} - \mathbf{v})$ 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 $\mathcal{Q}_m^n(\mathbf{a}) = \mathcal{Q}_m^{n|\leq L}(\mathbf{a}) + \mathcal{Q}_m^{n|>L}(\mathbf{a})$ 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 $u_{h_q;b}(\mathbf{w})$ also works nicely for the truncated functions $u_{h_q;b}(\mathbf{w})^{\leq L}$ and $u_{h_q;b}(\mathbf{w})^{>L}$. With regard to the limiting distributions, first note that $n^{-1/2} \mathcal{Q}_m^{n|>L}(\mathbf{a})$ 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:

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

Since $\ell r \asymp n$, it follows that $n^{-1/2} \mathcal{Q}_m^{n|>L}(\mathbf{a}) \Rightarrow 0$, so the limiting distributions of $n^{-1/2} \mathcal{Q}_m^n(\mathbf{a})$ and $n^{-1/2} \mathcal{Q}_m^{n|\leq L}(\mathbf{a})$ coincide when L is large enough.³² Next, observe that the random variable $|\eta_j^{\leq L}|$ obviously will have an upper bound, since the truncated polynomial $u_{h_q;b}(\mathbf{w})^{\leq L}$ will occur in the lower truncated version of eq. (B.67). Since the kernel function $K(\mathbf{w})$ by definition is bounded by some constant \mathcal{K} , it follows that $|\eta_j^{\leq L}|$ is bounded by

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

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

It remains to verify that the indicator functions $\mathbb{1}\left\{|\eta_j^{\leq L}| \geq \epsilon \sqrt{(\mathfrak{s}_\ell^2)^{\leq L}}\right\}$, from the lower truncated version of eq. (B.66), becomes zero when $n \rightarrow \infty$, which can be done by checking that the upper bound of $|\eta_j^{\leq L}|$ 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 $\sqrt{\ell rm}$, 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}_ℓ^2 share the same asymptotic expression, this becomes,

$$\frac{|\eta_j^{\leq L}|}{\sqrt{\ell rm}} \leq \mathcal{C} L \sqrt{\frac{mr}{\ell b_1 b_2}} \rightarrow 0, \quad \text{due to lemma C.3(c)}, \quad (\text{B.70a})$$

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

Assumption 2.2(b) ensures that $W_{h;b}$ (from definition B.10) converges to some non-zero matrix (as $h \rightarrow \infty$ and $\mathbf{b} \rightarrow \mathbf{0}^+$), and this implies that the limit of $\frac{1}{m} \sum_{h=1}^m \mathbf{a}'_h W_{h;b} \mathbf{a}_h$ in eq. (B.70b)

³²Truncation arguments often requires the threshold value L to go to ∞ 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).

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

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

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

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

The proof of item (a) is now complete, since the four random variables $n^{-1/2} \mathcal{Q}_{\bar{m}}^n(\mathbf{a})$, $n^{-1/2} \mathcal{Q}_{\bar{m}}^{n|\leq L}(\mathbf{a})$, $n^{-1/2} (S_n^{(1)})^{\leq L}$ 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. \square

Remark B.3. The statements in theorem B.20 has to be interpreted as an approximate asymptotic distributions valid for large m and n and small \mathbf{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 $K(\mathbf{w})$, which in the limit gives a degenerate Gaussian distribution in theorem B.20(b) (when $p > 1$). This degeneracy in itself would not have been any issue if the target of interest had been the asymptotic behaviour of $n^{-1/2} \mathcal{Q}_{\bar{m}}^n$, 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 $\hat{\boldsymbol{\theta}}_{\bar{m}:n}$, see appendix B.3 for details.

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

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

i.e. asymptotically mp -variate normal.

Proof. Lemma B.14 states that $\mathcal{Q}_{\bar{m}}^n$ and $\sqrt{b_1 b_2} \nabla_{\bar{m}} Q_{\bar{m}:n}(\boldsymbol{\theta}_{\bar{m}:b})$ have the same limiting distribution, and the result thus follows from theorem B.20(b). \square

B.3 The asymptotic results for $\hat{\boldsymbol{\theta}}_{\mathbf{v}|\bar{m}|b|p}$

The final details needed for the investigation of the asymptotic properties of $\hat{f}_{\mathbf{v}|p}^m(\omega)$ will now be presented, with a notation that discern between the two options $p = 1$ and $p = 5$. (Confer remark B.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 $\hat{\boldsymbol{\theta}}_{\mathbf{v}|\bar{m}|b|p}$ converges towards the true parameter vector $\boldsymbol{\theta}_{\mathbf{v}|\bar{m}|p}$ in the following manner.*

$$\sqrt{n(b_1 b_2)^{(p+1)/2}} \cdot \left(\hat{\boldsymbol{\theta}}_{\mathbf{v}|\bar{m}|b|p} - \boldsymbol{\theta}_{\mathbf{v}|\bar{m}|p}\right) \xrightarrow{d} N(\mathbf{0}, \Sigma_{\mathbf{v}|\bar{m}|p}), \quad (\text{B.74})$$

where $\Sigma_{\mathbf{v}|\bar{m}|p} := \bigoplus_{h=1}^m \Sigma_{\mathbf{v}|h|p}$, i.e. $\Sigma_{\mathbf{v}|\bar{m}|p}$ is the direct sum of the covariance matrices $\Sigma_{\mathbf{v}|h|p}$ that corresponds to $\sqrt{n(b_1 b_2)^{(p+1)/2}} \cdot \left(\hat{\boldsymbol{\theta}}_{\mathbf{v}|h|b|p} - \boldsymbol{\theta}_{\mathbf{v}|h|p} \right)$.

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 $Q_{\bar{m}:n}(\boldsymbol{\theta}_{\mathbf{v}|\bar{m}|b|p})$ in the general case where $m \rightarrow \infty$ and $\mathbf{b} \rightarrow \mathbf{0}^+$ when $n \rightarrow \infty$. 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 $\hat{\boldsymbol{\theta}}_{\mathbf{v}|\bar{m}|b|p}$ and the true parameter $\boldsymbol{\theta}_{\mathbf{v}|\bar{m}|p}$.

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

$$\sqrt{n} \cdot \left(\hat{\boldsymbol{\theta}}_{\mathbf{v}|\bar{m}|b|p} - \boldsymbol{\theta}_{\mathbf{v}|\bar{m}|p} \right) \xrightarrow{d} \mathbf{N}(\mathbf{0}, \Sigma_{\mathbf{v}|\bar{m}|p}), \quad (\text{B.75})$$

with $\Sigma_{\mathbf{v}|\bar{m}|p} := V_{\mathbf{v}|\bar{m}|p}^{-1} W_{\mathbf{v}|\bar{m}|p} V_{\mathbf{v}|\bar{m}|p}^{-1}$, where the $mp \times mp$ matrices $V_{\mathbf{v}|\bar{m}|p}$ and $W_{\mathbf{v}|\bar{m}|p}$ can be represented as

$$V_{\mathbf{v}|\bar{m}|p} = \bigoplus_{h=1}^m V_{\mathbf{v}|h|p}, \quad W_{\mathbf{v}|\bar{m}|p} = \bigoplus_{h=1}^m W_{\mathbf{v}|h|p}, \quad (\text{B.76})$$

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

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

For the general situation, when $m \rightarrow \infty$ and $\mathbf{b} \rightarrow \mathbf{0}^+$ when $n \rightarrow \infty$, 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 $K_{h:b}$ – and for the $p = 1$ case this is sufficient since the matrices $V_{\mathbf{v}|h|p}$ and $W_{\mathbf{v}|h|p}$ reduces to nonzero scalars.

For the $p = 5$ case, the limiting matrices of $V_{\mathbf{v}|h|p}$ and $W_{\mathbf{v}|h|p}$ turns out to have rank one, and additional scaling is thus 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 $\sqrt{(b_1 b_2)^3}$ when $p = 5$. \square

B.4 An extension to two different points, i.e. both \mathbf{v} and $\check{\mathbf{v}}$

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 $\hat{f}_{\mathbf{v}|p}^m(\omega)$ for a point \mathbf{v} that lies upon the diagonal (see theorem 2.7) or for general points $\mathbf{v} \in \mathbb{R}^2$ when the time series under investigation is time reversible (see theorem 2.8).

An investigation of the m -truncated estimates $\hat{f}_{\mathbf{v}|p}^m(\omega)$ for points $\mathbf{v} = (v_1, v_2)$ 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 $\mathbf{v} = (v_1, v_2)$ lies off the diagonal, and with the added requirement that the bivariate densities $g_h(\mathbf{y}_h)$ does not possess diagonal symmetry. With $\check{\mathbf{v}} = (v_2, v_1)$ the diagonal reflection of \mathbf{v} , the two parameter vectors $\hat{\boldsymbol{\theta}}_{\mathbf{v}|\bar{m}|b|p}$ and $\hat{\boldsymbol{\theta}}_{\check{\mathbf{v}}|\bar{m}|b|p}$ can be combined to a vector $\hat{\boldsymbol{\Theta}}_{\bar{m}|b|p}(\mathbf{v}, \check{\mathbf{v}}) = \left[\hat{\boldsymbol{\theta}}_{\mathbf{v}|\bar{m}|b|p}, \hat{\boldsymbol{\theta}}_{\check{\mathbf{v}}|\bar{m}|b|p} \right]'$, possessing the following asymptotic behaviour.

$$\sqrt{n(b_1 b_2)^{(p+1)/2}} \cdot \left(\hat{\boldsymbol{\Theta}}_{\bar{m}|b|p}(\mathbf{v}, \check{\mathbf{v}}) - \boldsymbol{\Theta}_{\bar{m}|p}(\mathbf{v}, \check{\mathbf{v}}) \right) \xrightarrow{d} \mathbf{N} \left(\mathbf{0}, \begin{bmatrix} \Sigma_{\mathbf{v}|\bar{m}|p} & 0 \\ 0 & \Sigma_{\check{\mathbf{v}}|\bar{m}|p} \end{bmatrix} \right), \quad (\text{B.77})$$

where the matrices $\Sigma_{\mathbf{v}|\bar{m}|p}$ and $\Sigma_{\check{\mathbf{v}}|\bar{m}|p}$ are as given in theorem B.22.

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

$$Q_{\bar{m}:n|p}(\boldsymbol{\Theta}_{\bar{m}|b|p}(\mathbf{v}, \check{\mathbf{v}})) := Q_{\bar{m}:n|p}(\boldsymbol{\theta}_{\mathbf{v}|\bar{m}|b|p}) + Q_{\bar{m}:n|p}(\boldsymbol{\theta}_{\check{\mathbf{v}}|\bar{m}|b|p}), \quad (\text{B.78})$$

i.e. the four requirements in items (A1) to (A4) of theorem B.1 must be verified for this new penalty function. The function $Q_{\bar{m}:n|p}$ on the right side of eq. (B.78) is the penalty function encountered in the investigation of $\boldsymbol{\theta}_{\mathbf{v}|\bar{m}|b|p}$, 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 \mathbf{v} in the first one and $\check{\mathbf{v}}$ in the second one.

The requirement that \mathbf{v} lies off the diagonal together with the requirement that none of the bivariate densities $g_h(\mathbf{y}_h)$ 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 $\boldsymbol{\theta}_{\mathbf{v}|\bar{m}|b|p}$ and $\boldsymbol{\theta}_{\check{\mathbf{v}}|\bar{m}|b|p}$. This implies that the arguments used to verify the three first requirements of theorem B.1 for the penalty function $Q_{\bar{m}:n|p}$ (see lemmas B.3 to B.5), also will work upon the combined penalty function $Q_{\bar{m}:n|p}$, and it will in particular be the case that the Hessian matrix $V_{\bar{m}|b:n|p}$ occurring in lemma B.4_p can be written as the direct sum of the matrices that corresponds to $Q_{\bar{m}:n|p}(\boldsymbol{\theta}_{\mathbf{v}|\bar{m}|b|p})$ and $Q_{\bar{m}:n|p}(\boldsymbol{\theta}_{\check{\mathbf{v}}|\bar{m}|b|p})$, i.e. $V_{\bar{m}|b|p}(\mathbf{v}, \check{\mathbf{v}}) = V_{\bar{m}|b:n|p}(\mathbf{v}) \oplus V_{\bar{m}|b:n|p}(\check{\mathbf{v}})$, where the points of interest has 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 $Q_{\bar{m}:n|p}$ requires some minor modifications of the constructions that was encountered in appendix B.2.1. Both $X_{hqt}^n(\mathbf{v})$ and $X_{hqt}^n(\check{\mathbf{v}})$ (for $h = 1, \dots, m$ and $q = 1, \dots, p$) are needed, and the final random variable will include both \mathbf{v} and $\check{\mathbf{v}}$ versions of the variables Z_{hqt}^n , \mathcal{Q}_{hqt}^n , $Z_{\bar{m}:t}^n(\mathbf{a})$, $Z_{\bar{m}:t}^n$, $\mathcal{Q}_{\bar{m}}^n(\mathbf{a})$ and $\mathcal{Q}_{\bar{m}}^n$.

A minor revision of lemma B.14 proves that the same limiting distribution occurs for the $\sqrt{b_1 b_2}$ -scaled gradient of $Q_{\bar{m}:n|p}(\boldsymbol{\Theta}_{\bar{m}|b|p}(\mathbf{v}, \check{\mathbf{v}}))$ and for the random variable $\mathcal{Q}_{\bar{m}}^n(\mathbf{v}, \check{\mathbf{v}}) := [\mathcal{Q}_{\bar{m}}^n(\mathbf{v})', \mathcal{Q}_{\bar{m}}^n(\check{\mathbf{v}})']'$, and it is easy to see that $Z_{\bar{m}:t}^n(\mathbf{a}_1, \mathbf{a}_2; \mathbf{v}, \check{\mathbf{v}}) := Z_{\bar{m}:t}^n(\mathbf{a}_1; \mathbf{v}) + Z_{\bar{m}:t}^n(\mathbf{a}_2; \check{\mathbf{v}})$ must take the place of $Z_{\bar{m}:t}^n(\mathbf{a})$ in the existing proofs. The key ingredient for the asymptotic investigation of $Z_{\bar{m}:t}^n(\mathbf{a}_1, \mathbf{a}_2; \mathbf{v}, \check{\mathbf{v}})$ is a simple extension of lemma B.15(c) such that it also covers the ‘cross-term’ cases $E[X_{hqt}^n(\mathbf{v}) \cdot X_{jrk}^n(\check{\mathbf{v}})]$ and verifies that these cases are asymptotically negligible. This follows from the results stated in lemma C.7

The statement for $Z_{\bar{m}:t}^n(\mathbf{a})$ 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 $Z_{\bar{m}:t}^n(\mathbf{a}_1, \mathbf{a}_2; \mathbf{v}, \check{\mathbf{v}})$, it follows that

$$\begin{aligned} Z_{\bar{m}:i}^n(\mathbf{a}_1, \mathbf{a}_2; \mathbf{v}, \check{\mathbf{v}}) \cdot Z_{\bar{m}:k}^n(\mathbf{a}_1, \mathbf{a}_2; \mathbf{v}, \check{\mathbf{v}}) &= Z_{\bar{m}:i}^n(\mathbf{a}_1; \mathbf{v}) \cdot Z_{\bar{m}:k}^n(\mathbf{a}_1; \mathbf{v}) + Z_{\bar{m}:i}^n(\mathbf{a}_1; \mathbf{v}) \cdot Z_{\bar{m}:k}^n(\mathbf{a}_2; \check{\mathbf{v}}) \\ &\quad + Z_{\bar{m}:k}^n(\mathbf{a}_1; \mathbf{v}) \cdot Z_{\bar{m}:i}^n(\mathbf{a}_2; \check{\mathbf{v}}) + Z_{\bar{m}:i}^n(\mathbf{a}_2; \check{\mathbf{v}}) \cdot Z_{\bar{m}:k}^n(\mathbf{a}_2; \check{\mathbf{v}}), \end{aligned} \quad (\text{B.79})$$

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

$$\frac{1}{n} \sum_{\substack{i,k=1 \\ i \neq k}}^n |Z_{\bar{m}:i}^n(\mathbf{a}_1; \mathbf{v}) \cdot Z_{\bar{m}:k}^n(\mathbf{a}_2; \check{\mathbf{v}})| = o(1), \quad (\text{B.80})$$

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

$$\sum_{\ell=1}^{n-1} \left(1 - \frac{\ell}{n}\right) I_{\bar{m}:\ell}^n(\mathbf{a}_1, \mathbf{a}_2; \mathbf{v}, \check{\mathbf{v}}) + \sum_{\ell=1}^{n-1} \left(1 - \frac{\ell}{n}\right) I_{\bar{m}:\ell}^n(\mathbf{a}_2, \mathbf{a}_1; \check{\mathbf{v}}, \mathbf{v}), \quad (\text{B.81})$$

where $I_{\bar{m}:\ell}^n(\mathbf{a}_1, \mathbf{a}_2; \mathbf{v}, \check{\mathbf{v}}) := |\mathbb{E}[Z_{\bar{m}:0}^n(\mathbf{a}_1, \mathbf{v}) \cdot Z_{\bar{m}:\ell}^n(\mathbf{a}_2, \check{\mathbf{v}})]|$, with $I_{\bar{m}:\ell}^n(\mathbf{a}_2, \mathbf{a}_1; \check{\mathbf{v}}, \mathbf{v})$ 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 $Z_{\bar{m}:t}^n(\mathbf{a}_1, \mathbf{a}_2; \mathbf{v}, \check{\mathbf{v}})$ is straight forward, i.e. the standard formula for the variance of a sum of random variables gives

$$\text{Var}(Z_{\bar{m}:t}^n(\mathbf{a}_1, \mathbf{a}_2; \mathbf{v}, \check{\mathbf{v}})) = \text{Var}(Z_{\bar{m}:t}^n(\mathbf{a}_1, \mathbf{v})) + 2 \text{Cov}(Z_{\bar{m}:t}^n(\mathbf{a}_1, \mathbf{v}), Z_{\bar{m}:t}^n(\mathbf{a}_2, \check{\mathbf{v}})) + \text{Var}(Z_{\bar{m}:t}^n(\mathbf{a}_2, \check{\mathbf{v}})),$$

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

$$\mathbf{a}'_{\bar{m}} \cdot W_{\bar{m}:b} \cdot \mathbf{a}_{\bar{m}} := [\mathbf{a}'_1, \mathbf{a}'_2] \cdot (W_{\bar{m}:b}(\mathbf{v}) \oplus W_{\bar{m}:b}(\check{\mathbf{v}})) \cdot \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{bmatrix} = \mathbf{a}'_1 \cdot W_{\bar{m}:b}(\mathbf{v}) \cdot \mathbf{a}_1 + \mathbf{a}'_2 \cdot W_{\bar{m}:b}(\check{\mathbf{v}}) \cdot \mathbf{a}_2, \quad (\text{B.82})$$

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 $W_{\bar{m}:b} := W_{\bar{m}:b}(\mathbf{v}) \oplus W_{\bar{m}:b}(\check{\mathbf{v}})$, 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

$$n^{-1/2} Q_{\bar{m}:n|p}(\Theta_{\bar{m}|b|p}(\mathbf{v}, \check{\mathbf{v}})) \xrightarrow{d} \mathbf{N}(\mathbf{0}, W_{\bar{m}:b}(\mathbf{v}) \oplus W_{\bar{m}:b}(\check{\mathbf{v}})). \quad (\text{B.83})$$

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. \square

Remark B.4. 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 $\widehat{\boldsymbol{\theta}}_{\mathbf{v}_i|\bar{m}|b|p}$ corresponding to different points $\{\mathbf{v}_i\}_{i=1}^\nu$ will be jointly asymptotically normal and pairwise asymptotically independent. The asymptotically independent property are inherited by the corresponding estimated local Gaussian spectral densities $\widehat{f}_{\mathbf{v}_i|p}^m(\omega)$, and this enables an alternative smoothing strategy for the estimated local Gaussian spectral densities at a given point \mathbf{v} , see section 4.4. However, the added computational cost incurred by such an estimation approach may make this a less interesting topic of investigation.

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 $\rho_{\mathbf{v}|p}(h)$ 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 \mathbf{b} , whereas appendix C.4 shows that the assumptions on the kernel function $K(\mathbf{w})$ and the score functions $u_{h\mathbf{q};\mathbf{b}}(\mathbf{w})$ implies that some integrals are finite (which implies that assumption 2.1(g) will be trivially satisfied if the bivariate densities $g_h(\mathbf{y}_h)$ are finite). Appendix C.5 contains a few basic definitions/comments related to α -mixing, σ -algebras and L^ν -spaces, and finally appendix C.6 presents a comparison of the *five-parameter* versus the *one-parameter* local Gaussian approximation, in order to pinpoint why $\widehat{f}_{\mathbf{v}|5}^m(\omega)$ should be used instead of $\widehat{f}_{\mathbf{v}|1}^m(\omega)$.

C.1 The diagonal folding property of $\rho_{\mathbf{v}|p}(h)$

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

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

$$\rho_{\mathbf{v}|p}(-h) = \rho_{\check{\mathbf{v}}|p}(h), \quad (\text{C.1})$$

where $\check{\mathbf{v}} = (v_2, v_1)$ is the diagonal reflection of \mathbf{v} .

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

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

gives the following property³³ for the probability density functions,

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

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

Remark C.1. The $p = 1$ case corresponds to the situation where it from the outset has been assumed that $\mu_1 = \mu_2 = 0$ and $\sigma_{11} = \sigma_{22} = 1$.

Remark C.2. 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 i.e. implies that the local Gaussian spectral densities $f_{v|p}(\omega)$ always are real-valued for such time series, see definition 2.2 and theorem 2.8.

The following simple result follows immediately from definition 2.2.

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

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

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

$$\rho_{v|p}(-h) = \rho_{v|p}(h). \quad (\text{C.5})$$

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.

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

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. \square

C.3 Two limit theorems

This section contains two lemmas. 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.

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

³³This 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.

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

(b) There exists integers s and constants $c := c_n \rightarrow \infty$, such that

$$c \cdot s = o\left(\sqrt{nb_1b_2/m}\right), \quad \sqrt{nm/b_1b_2} \cdot c \cdot \alpha(s - m + 1) \rightarrow 0. \quad (\text{C.6})$$

(c) There exists integers s and constants c , such that with r , ℓ and ϑ given as the integers

$$r = r_n := \left\lfloor \frac{\sqrt{nb_1b_2/m}}{c} \right\rfloor, \quad \ell = \ell_n := \left\lfloor \frac{n}{r + s} \right\rfloor, \quad \vartheta = \vartheta_n := s - m + 1, \quad (\text{C.7})$$

the following limits occur when $n \rightarrow \infty$:

$$\frac{s}{r} \rightarrow 0; \quad \ell\alpha(\vartheta) \rightarrow 0; \quad \frac{mr}{n} \rightarrow 0; \quad \frac{mr}{\ell b_1 b_2} \rightarrow 0; \quad \frac{m\ell s}{n} \rightarrow 0. \quad (\text{C.8})$$

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 $m = o((nb_1b_2)^\xi)$, for any $\xi \in (0, \frac{1}{3})$, 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 $s = o(\sqrt{nb_1b_2/m})$ 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

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

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

$$\begin{aligned} \frac{s}{\sqrt{nb_1b_2/m}} &\asymp \frac{m \cdot (nb_1b_2)^{\zeta/2}}{(nb_1b_2/m)^{1/2}} = \frac{m^{3/2}}{(nb_1b_2)^{(1-\zeta)/2}} = \left[\frac{m}{(nb_1b_2)^{(1-\zeta)/3}} \right]^{3/2} \\ &= \left[\frac{1}{(nb_1b_2)^{2\zeta/3}} \cdot \frac{m}{(nb_1b_2)^{1/3-\zeta}} \right]^{3/2} \rightarrow \left[\frac{1}{\infty} \cdot 0 \right]^{3/2} = 0. \end{aligned} \quad (\text{C.10})$$

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

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

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

Since $1 - \eta$ and $\eta/2$ are in $(0, 1)$, it follows from assumption 2.3(b) that s and c goes to ∞ as required. A quick inspection reveals that the product $c \cdot s$ is $o(\sqrt{nb_1b_2/m})$, 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

η 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 $x \rightarrow \infty$, 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 ∞ . With these observations, all except the last limit in eq. (C.8) are trivial to prove, i.e.

$$\frac{s}{r} \asymp \frac{s}{\frac{\sqrt{nb_1 b_2/m}}{c}} = \frac{c \cdot s}{\sqrt{nb_1 b_2/m}} \rightarrow 0, \quad (\text{C.12a})$$

$$\ell\alpha(\vartheta) \leq \frac{n}{r+s} \alpha(\vartheta) \asymp \frac{n}{r} \alpha(\vartheta) \asymp \frac{n}{\frac{\sqrt{nb_1 b_2/m}}{c}} \alpha(\vartheta) = \sqrt{nm/b_1 b_2} \cdot c \cdot \alpha(\vartheta) \rightarrow 0, \quad (\text{C.12b})$$

$$\frac{mr}{n} \leq \frac{\frac{\sqrt{nb_1 b_2/m}}{c}}{n/m} = \frac{\sqrt{b_1 b_2}}{c\sqrt{n/m}} \rightarrow \frac{0}{\infty \cdot \infty} = 0, \quad (\text{C.12c})$$

$$\frac{mr}{\ell b_1 b_2} \asymp \frac{mr}{\frac{n}{r+s} b_1 b_2} = \frac{r(r+s)}{nb_1 b_2/m} \asymp \frac{r^2}{nb_1 b_2/m} \leq \frac{\frac{nb_1 b_2/m}{c^2}}{nb_1 b_2/m} = \frac{1}{c^2} \rightarrow 0. \quad (\text{C.12d})$$

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

$$\begin{aligned} \frac{m\ell s}{n} &\leq \frac{m \frac{n}{r+s} s}{n} = m \frac{s}{r+s} \asymp m \frac{s}{r} \asymp m \frac{c \cdot s}{\sqrt{nb_1 b_2/m}} \leq m \frac{\left(\sqrt{nb_1 b_2/m}\right)^{1-\eta/2}}{\sqrt{nb_1 b_2/m}} \\ &= \frac{m}{(nb_1 b_2/m)^{\eta/4}} = \frac{m^{1+\eta/4}}{(nb_1 b_2)^{\eta/4}} = \left(\frac{m}{(nb_1 b_2)^{\eta/(4+\eta)}}\right)^{(4+\eta)/4}. \end{aligned} \quad (\text{C.13})$$

Assumption 2.3(e) states that $m = o\left((nb_1 b_2)^{\tau/(2+5\tau)-\lambda}\right)$, and it is consequently sufficient to show that an η can be found which gives $\tau/(2+5\tau) - \lambda \leq p(\eta) := \eta/(4+\eta)$. 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 η that satisfies $\tau/(2+5\tau) - \lambda \leq p(\eta) < \tau/(2+5\tau)$, which concludes the proof. \square

Lemma C.4. *Under assumption 2.3, the sequence of integers defined by $k_n + 1 := \lceil m^{2/a} \cdot |b_1 b_2|^{(2-\nu)/a\nu} \rceil$ satisfies the following two limit requirements.*

- (a) $k_n \rightarrow \infty$.
- (b) $k_n m^2 b_1 b_2 \rightarrow 0$.

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 $m \rightarrow \infty$ and $\mathbf{b} \rightarrow \mathbf{0}^+$ when $n \rightarrow \infty$, it follows that $k_n \rightarrow \infty$, which proves item (a).

For item (b), observe that $k_n = \lceil m^{2/a} \cdot |b_1 b_2|^{(2-\nu)/a\nu} \rceil - 1 < m^{2/a} \cdot |b_1 b_2|^{(2-\nu)/a\nu}$ implies

$$k_n m^2 b_1 b_2 < (m^{2/a} \cdot |b_1 b_2|^{(2-\nu)/a\nu}) \cdot m^2 b_1 b_2 \quad (\text{C.14a})$$

$$= m^{2(1+1/a)} \cdot |b_1 b_2|^{1+(2-\nu)/a\nu} \quad (\text{C.14b})$$

$$\leq m^{2(1+1/a)} \cdot |(b_1 \vee b_2)^2|^{1+(2-\nu)/a\nu} \quad (\text{C.14c})$$

$$= \{m^{\{1+1/a\}/\{1+(2-\nu)/a\nu\}} \cdot (b_1 \vee b_2)\}^{2(1+(2-\nu)/a\nu)} \quad (\text{C.14d})$$

$$= \{m^{\{\nu(a+1)\}/\{\nu(a-1)+2\}} \cdot (b_1 \vee b_2)\}^{2(1+(2-\nu)/a\nu)}. \quad (\text{C.14e})$$

An inspection of the outermost exponent reveals

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

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

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 $K(\mathbf{w})$ and $u_{h_q; \mathbf{b}}(\mathbf{w})$. Note that the treatment in this section exploits the property that the functions $u_{h_q; \mathbf{b}}(\mathbf{w})$ 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 $K(\mathbf{w})$ from definition B.9 (page 51), and $\nu > 2$ from assumption 2.1(b) (page 12), the following holds:*

$$(a) \left| \int_{\mathbb{R}^2} K(w_1, w_2) w_1^k w_2^\ell dw_1 dw_2 \right| < \infty, \quad k, \ell \geq 0 \text{ and } k + \ell \leq 5.$$

$$(b) \left| \int_{\mathbb{R}^2} K(w_1, w_2)^2 w_1^k w_2^\ell dw_1 dw_2 \right| < \infty, \quad k, \ell \geq 0 \text{ and } k + \ell \leq 5.$$

$$(c) K(w_1, w_2) w_1^k w_2^\ell \in L^\nu, \quad k, \ell \geq 0 \text{ and } k + \ell \leq 2.$$

Proof. Since the kernel function by definition is non-negative, it follows that

$$\left| \int_{\mathbb{R}^2} K(w_1, w_2) w_1^k w_2^\ell dw_1 dw_2 \right| \leq \int_{\mathbb{R}^2} K(w_1, w_2) |w_1^k w_2^\ell| dw_1 dw_2, \quad (\text{C.16})$$

which proves item (a), since eq. (B.27d) of definition B.9 implies that this is finite for the specified range of k and ℓ .

Since the kernel function is bounded, there is some constant \mathcal{C} such that $K(\mathbf{w}) \leq \mathcal{C}$, which implies that

$$\left| \int_{\mathbb{R}^2} K(w_1, w_2)^2 w_1^k w_2^\ell dw_1 dw_2 \right| \leq \mathcal{C} \left| \int_{\mathbb{R}^2} K(w_1, w_2) w_1^k w_2^\ell dw_1 dw_2 \right|, \quad (\text{C.17})$$

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,

$$\left(\int_{\mathbb{R}^2} |K(w_1, w_2) w_1^k w_2^\ell|^\nu \mathrm{d}w_1 \mathrm{d}w_2 \right)^{1/\nu} \leq \mathcal{C}^{(\nu-1)/\nu} \left(\int_{\mathbb{R}^2} K(w_1, w_2) |w_1^k w_2^\ell|^\nu \mathrm{d}w_1 \mathrm{d}w_2 \right)^{1/\nu}, \quad (\text{C.18})$$

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 $\mathcal{A}_{k\ell} = \{\mathbf{w} : |w_1^k w_2^\ell| \leq 1\}$ and $\mathcal{A}_{k\ell}^c = \mathbb{R}^2 \setminus \mathcal{A}_{k\ell}$, it follows from the non-negativeness of $K(\mathbf{w})$, and eqs. (B.27a) and (B.27d) of definition B.9, that

$$\int_{\mathcal{A}_{k\ell}} K(w_1, w_2) |w_1^k w_2^\ell|^\nu \mathrm{d}w_1 \mathrm{d}w_2 \leq \int_{\mathcal{A}_{k\ell}} K(w_1, w_2) \mathrm{d}w_1 \mathrm{d}w_2 \leq \int_{\mathbb{R}^2} K(w_1, w_2) \mathrm{d}w_1 \mathrm{d}w_2 = 1, \quad (\text{C.19a})$$

$$\begin{aligned} \int_{\mathcal{A}_{k\ell}^c} K(w_1, w_2) |w_1^k w_2^\ell|^\nu \mathrm{d}w_1 \mathrm{d}w_2 &\leq \int_{\mathcal{A}_{k\ell}^c} K(w_1, w_2) |w_1^k w_2^\ell|^{\lceil \nu \rceil} \mathrm{d}w_1 \mathrm{d}w_2 \\ &\leq \int_{\mathbb{R}^2} K(w_1, w_2) |w_1^{k\lceil \nu \rceil} w_2^{\ell\lceil \nu \rceil}| \mathrm{d}w_1 \mathrm{d}w_2 < \infty, \end{aligned} \quad (\text{C.19b})$$

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), $K(w_1, w_2) w_1^k w_2^\ell \in L^\nu$. \square

Lemma C.6. *The following holds for $u_{h,q;b}(\mathbf{w})$ and $K_{h;b}(\mathbf{y}_h - \mathbf{v})$ from definitions B.7 and B.9, and $\nu > 2$ from assumption 2.1(b):*

- (a) $\int_{\mathbb{R}^2} \sqrt{b_1 b_2} K_{h;b}(\boldsymbol{\zeta} - \mathbf{v}) u_{h,q;b}(\boldsymbol{\zeta}) \mathrm{d}\boldsymbol{\zeta} = O(\sqrt{b_1 b_2})$.
- (b) $\left(\int_{\mathbb{R}^2} |\sqrt{b_1 b_2} K_{h;b}(\boldsymbol{\zeta} - \mathbf{v}) u_{h,q;b}(\boldsymbol{\zeta})|^\nu \mathrm{d}\boldsymbol{\zeta} \right)^{1/\nu} = O(|b_1 b_2|^{(2-\nu)/2\nu})$.
- (c) Let $\mathcal{K}_{qr,hj;b}(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2) := K_{h;b}(\boldsymbol{\zeta}_1 - \mathbf{v}) K_{j;b}(\boldsymbol{\zeta}_2 - \mathbf{v}) u_{h,q;b}(\boldsymbol{\zeta}_1) u_{j,r;b}(\boldsymbol{\zeta}_2)$, where $\boldsymbol{\zeta}_1$ and $\boldsymbol{\zeta}_2$ either coincide completely (bivariate), have one common component (trivariate), or have no common components (tetrivariate). Let κ be the number of variates, and let $\mathrm{d}\boldsymbol{\zeta}(\kappa)$ represent the corresponding κ -variate differential. Then,

$$\int_{\mathbb{R}^\kappa} (b_1 b_2) \mathcal{K}_{qr,hj;b}(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2) \mathrm{d}\boldsymbol{\zeta}(\kappa) = \begin{cases} u_{h,q;b}(\mathbf{v}) u_{j,r;b}(\mathbf{v}) \int_{\mathbb{R}^2} K(\mathbf{w})^2 \mathrm{d}\mathbf{w} + O(b_1 \vee b_2) & \kappa = 2, \\ O(b_1 \wedge b_2) & \kappa = 3, \\ O(b_1 b_2) & \kappa = 4. \end{cases}$$

Proof. Recalling the definition of $K_{h;b}(\mathbf{y}_h - \mathbf{v})$ from eq. (B.28), the integral in item (a) can be written as

$$\int_{\mathbb{R}^2} \sqrt{b_1 b_2} \cdot \frac{1}{b_1 b_2} K\left(\frac{\zeta_1 - v_1}{b_1}, \frac{\zeta_2 - v_2}{b_2}\right) u_{h,q;b}(\zeta_1, \zeta_2) \mathrm{d}\zeta_1 \mathrm{d}\zeta_2, \quad (\text{C.20})$$

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

$$\begin{aligned} & \int_{\mathbb{R}^2} \frac{\sqrt{b_1 b_2}}{b_1 b_2} K(w_1, w_2) u_{h_q; \mathbf{b}}(b_1 w_1 + v_1, b_2 w_2 + v_2) (b_1 dw_1) (b_2 dw_2) \\ &= \sqrt{b_1 b_2} \cdot \int_{\mathbb{R}^2} K(w_1, w_2) u_{h_q; \mathbf{b}}(b_1 w_1 + v_1, b_2 w_2 + v_2) dw_1 dw_2. \end{aligned} \quad (\text{C.21})$$

Since $u_{h_q; \mathbf{b}}(\mathbf{w})$ is a bivariate polynomial, it is clear that $u_{h_q; \mathbf{b}}(b_1 w_1 + v_1, b_2 w_2 + v_2)$ can be written as

$$u_{h_q; \mathbf{b}}(v_1, v_2) + b_1 c_1 w_1 + b_2 c_2 w_2 + b_1^2 c_{11} w_1^2 + b_1 b_2 c_{12} w_1 w_2 + b_2^2 c_{22} w_2^2, \quad (\text{C.22})$$

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 $O(\sqrt{b_1 b_2})$ when $\mathbf{b} \rightarrow \mathbf{0}^+$, and the conclusion of item (a) follows.

The substitution used in item (a) can also be applied for item (b), resulting in

$$\begin{aligned} & \left(\int_{\mathbb{R}^2} \left| \sqrt{b_1 b_2} \cdot \frac{1}{b_1 b_2} K(w_1, w_2) u_{h_q; \mathbf{b}}(b_1 w_1 + v_1, b_2 w_2 + v_2) \right|^\nu (b_1 dw_1) (b_2 dw_2) \right)^{1/\nu} \\ &= |b_1 b_2|^{(2-\nu)/2\nu} \left(\int_{\mathbb{R}^2} |K(w_1, w_2) u_{h_q; \mathbf{b}}(b_1 w_1 + v_1, b_2 w_2 + v_2)|^\nu dw_1 dw_2 \right)^{1/\nu}. \end{aligned} \quad (\text{C.23})$$

Note that this represent the norm in L^ν -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

$$|b_1 b_2|^{(2-\nu)/2\nu} \left(u_{h_q; \mathbf{b}}(v_1, v_2) + b_1 e_1 w_1 + b_2 e_2 w_2 + b_1^2 e_{11} w_1^2 + b_1 b_2 e_{12} w_1 w_2 + b_2^2 e_{22} w_2^2 \right), \quad (\text{C.24})$$

which is dominated by the $|b_1 b_2|^{(2-\nu)/2\nu}$ -term when $\mathbf{b} \rightarrow \mathbf{0}^+$, as stated in item (b).

The investigation of item (c) requires different substitutions depending on the κ for the configuration under investigation. Noting that the integrand in addition to the scaling factor $b_1 b_2$ always contains the product $K_{h; \mathbf{b}}(\zeta_1 - \mathbf{v}) K_{j; \mathbf{b}}(\zeta_2 - \mathbf{v})$, it follows that it regardless of the value of κ 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 $b_1 b_2 dw_1 dw_2$ when $\kappa = 2$, $b_1^2 b_2 dw_1 dw_2 dw_3$ or $b_1 b_2^2 dw_1 dw_2 dw_3$ when $\kappa = 3$, and $b_1^2 b_2^2 dw_1 dw_2 dw_3 dw_4$ when $\kappa = 4$.

For the bivariate case, the substitution from item (a) gives an expression of the following form,

$$\int_{\mathbb{R}^2} K(w_1, w_2)^2 \cdot \mathcal{U}(w_1, w_2) dw_1 dw_2, \quad (\text{C.25})$$

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 $u_{h_q; \mathbf{b}}(\mathbf{v}) u_{j_r; \mathbf{b}}(\mathbf{v})$. From lemma C.6(b) it follows that this will be a finite integral, and as $\mathbf{b} \rightarrow \mathbf{0}^+$ the result will be as given for the $\kappa = 2$ case of item (c).

For the trivariate case, the overlap between ζ_1 and ζ_2 will belong to one of the following configurations, (i) $\zeta_1 = (\zeta_1, \zeta_2)$ and $\zeta_2 = (\zeta_1, \zeta_3)$, (ii) $\zeta_1 = (\zeta_1, \zeta_2)$ and $\zeta_2 = (\zeta_3, \zeta_1)$, (iii) $\zeta_1 = (\zeta_1, \zeta_2)$ and $\zeta_2 = (\zeta_2, \zeta_3)$, or (iv) $\zeta_1 = (\zeta_1, \zeta_2)$ and $\zeta_2 = (\zeta_3, \zeta_2)$. 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,

$$K((\zeta_1 - v_1)/b_1, (\zeta_2 - v_2)/b_2) \cdot K((\zeta_2 - v_1)/b_1, (\zeta_3 - v_2)/b_2). \quad (\text{C.26})$$

When the substitution

$$w_1 = (\zeta_1 - v_1)/b_1, \quad w_2 = (\zeta_2 - v_2)/b_2, \quad w_3 = (\zeta_3 - v_2)/b_2, \quad (\text{C.27})$$

is used, the following component occurs in the transformed integrand,

$$\mathcal{K}(w_1, w_2, w_3) := K(w_1, w_2) \cdot K([(b_2 w_2 + v_2) - v_1]/b_1, w_3). \quad (\text{C.28})$$

The argument $[(b_2 w_2 + v_2) - v_1]/b_1$ 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 $\ell \in \{0, 1, 2\}$,

$$\int_{\mathbb{R}^1} \mathcal{K}(w_1, w_2, w_3) w_3^\ell \mathbf{d}w_3 = K(w_1, w_2) \cdot \int_{\mathbb{R}^1} K([(b_2 w_2 + v_2) - v_1]/b_1, w_3) w_3^\ell \mathbf{d}w_3 \quad (\text{C.29a})$$

$$= K(w_1, w_2) \cdot \mathcal{K}_{2; \ell}([(b_2 w_2 + v_2) - v_1]/b_1) \quad (\text{C.29b})$$

$$\leq \mathcal{D}_{2; \ell} \cdot K(w_1, w_2), \quad (\text{C.29c})$$

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

$$b_2 \int_{\mathbb{R}^3} \mathcal{K}(w_1, w_2, w_3) \cdot \mathcal{U}(w_1, w_2, w_3) \mathbf{d}w_1 \mathbf{d}w_2 \mathbf{d}w_3, \quad (\text{C.30})$$

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 $\mathbf{b} \rightarrow \mathbf{0}^+$. 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 $\mathbf{b} \rightarrow \mathbf{0}^+$ – and from this it 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 ζ_1 and ζ_2 implies that the tetravariate integral, after the obvious substitution, corresponds to an expression of the form

$$b_1 b_2 \left(\int_{\mathbb{R}^2} K(\mathbf{w}) u_{h_q; \mathbf{b}}(\zeta(\mathbf{w})) \, d\mathbf{w} \right) \cdot \left(\int_{\mathbb{R}^2} K(\mathbf{w}) u_{j_r; \mathbf{b}}(\zeta(\mathbf{w})) \, d\mathbf{w} \right), \quad (\text{C.31})$$

where $\zeta(\mathbf{w}) = (b_1 w_1 + v_1, b_2 w_2 + v_2)$. 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 $\mathbf{b} \rightarrow \mathbf{0}^+$, which concludes the proof of item (c). \square

Note that the bivariate case of lemma C.6(c) only considers the configuration where the components of ζ_1 and ζ_2 coincide completely, while the configuration where $\zeta_1 = (\zeta_1, \zeta_2)$ and ζ_2 is the diagonal reflection (ζ_2, ζ_1) has been left out. This restriction does not pose a problem for the asymptotic investigation of $\hat{f}_{v|p}^m(\omega)$ when the point $\mathbf{v} = (v_1, v_2)$ 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 \mathbf{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 \mathbf{v} and the other use the diagonally reflected point $\check{\mathbf{v}} = (v_2, v_1)$.

Lemma C.7. *The following holds for $u_{h_q; \mathbf{b}}(\mathbf{w})$ and $K_{h; \mathbf{b}}(\mathbf{y}_h - \mathbf{v})$ from definitions B.7 and B.9, when the point $\mathbf{v} = (v_1, v_2)$ does not coincide with its diagonal reflection $\check{\mathbf{v}} = (v_2, v_1)$, i.e. $v_1 \neq v_2$.*

Let $\mathcal{K}_{q_r, h_j; \mathbf{b}}(\zeta_1, \zeta_2; \mathbf{v}, \check{\mathbf{v}}) := K_{h; \mathbf{b}}(\zeta_1 - \mathbf{v}) K_{j; \mathbf{b}}(\zeta_2 - \check{\mathbf{v}}) u_{h_q; \mathbf{b}}(\zeta_1) u_{j_r; \mathbf{b}}(\zeta_2)$, where ζ_1 and ζ_2 either are diagonal reflections of each other (bivariate), have one common component (trivariate), or have no common components (tetravariate). Let κ be the number of variates, and let $d\zeta(\kappa)$ represent the corresponding κ -variate differential. Then,

$$\int_{\mathbb{R}^\kappa} (b_1 b_2) \mathcal{K}_{q_r, h_j; \mathbf{b}}(\zeta_1, \zeta_2; \mathbf{v}, \check{\mathbf{v}}) \, d\zeta(\kappa) = \begin{cases} o(1) & \kappa = 2, \\ O(b_1 \wedge b_2) & \kappa = 3, \\ O(b_1 b_2) & \kappa = 4. \end{cases}$$

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 $\int_{\mathbb{R}^2} K(w_1, w_2)^2 \cdot \mathcal{U}(w_1, w_2) \, dw_1 \, dw_2$ from eq. (C.25) is replaced with a sum of integrals of the form,

$$\int_{\mathbb{R}^2} K(w_1 + (v_1 - v_2) / b_1, w_2 + (v_2 - v_1) / b_2) \cdot K(w_1, w_2) w_1^k w_2^\ell \, dw_1 \, dw_2, \quad (\text{C.32})$$

where $k, \ell \geq 0$ and $k + \ell \leq 4$. and the integrands of these integrals goes to zero when $\mathbf{b} \rightarrow \mathbf{0}^+$, 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 $K(w_1, w_2)w_1^k w_2^l$, and together this implies that the integrand in eq. (C.32) must go to zero when $\mathbf{b} \rightarrow \mathbf{0}^+$, and the integral thus becomes asymptotically negligible. \square

Remark C.3. 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 $K(w)$ 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 $K_{h;b}(\zeta_1 - v)$ and $K_{j;b}(\zeta_2 - v)$ 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 σ -algebras, α -mixing and L^ν -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 σ -algebras

The σ -algebras related to the process $\{Y_t\}_{t \in \mathbb{Z}}$, will be denoted

$$\mathcal{F}_t^s := \sigma(Y_t, \dots, Y_s), \tag{C.33}$$

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 $\xi(\mathbf{y}_m)$ from \mathbb{R}^{m+1} to \mathbb{R} , i.e. $\mathcal{Y}_{m:t} := \xi(\mathbf{Y}_{m:t})$, then $\mathcal{Y}_{m:t} \in \mathcal{F}_t^{t+m}$.

Inheritance of α -mixing

The coefficients in the strong mixing property mentioned in assumption 2.1(b), is given by

$$\alpha(s | Y_t) := \sup \{ |P(A \cap B) - P(A)P(B)| : -\infty < t < \infty, A \in \mathcal{F}_{-\infty}^t, B \in \mathcal{F}_{t+s}^\infty \}, \tag{C.34}$$

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 α -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. (2.17) will be inherited by the process $\mathcal{Y}_{m:t}$, i.e. with ν and a as introduced in assumption 2.1(b), the following holds true

$$\sum_{j=1}^{\infty} j^a [\alpha(j | \mathcal{Y}_{m:t})]^{1-2/\nu} < \infty. \tag{C.36}$$

Related L^ν -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^ν -spaces outlined below.³⁴

³⁴These definitions are normally presented with p used instead of ν .

First of all, when a measure space $(\Omega, \mathcal{G}, \mu)$ is given, then for $1 \leq \nu < \infty$, the space $L^\nu := L^\nu(\Omega, \mathcal{G}, \mu)$ is defined to be the class of measurable real functions ζ for which $|\zeta|^\nu$ is integrable, that is,

$$\zeta(z) \in L^\nu \iff \int_{\Omega} |\zeta(z)|^\nu d\mu < \infty. \quad (\text{C.37})$$

The L^ν -spaces related to the processes $\mathbf{Y}_{h:t}$ and $\mathbf{Y}_{\bar{m}:t}$ will henceforth be denoted by

$$L_h^\nu \text{ — the } L^\nu \text{ spaces related to the densities } g_h, \quad (\text{C.38a})$$

$$L_{\bar{m}}^\nu \text{ — the } L^\nu \text{ space related to the density } g_{\bar{m}}. \quad (\text{C.38b})$$

These L^ν 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 ν -norm defined by

$$\|\zeta(z)\|_\nu := \left(\int_{\Omega} |\zeta(z)|^\nu d\mu \right)^{1/\nu} = (\mathbb{E}[|\zeta(Z)|^\nu])^{1/\nu} \quad (\text{C.39})$$

and the Minkowski's inequality (i.e. the triangle inequality for L^ν -spaces) will play a central role in the investigation later on,

$$\|\zeta_1(z) + \zeta_2(z)\|_\nu \leq \|\zeta_1(z)\|_\nu + \|\zeta_2(z)\|_\nu. \quad (\text{C.40})$$

The main reason for the introduction of these L^ν -spaces are the following observation: With Z a random variable on $(\Omega, \mathcal{G}, \mu)$, the definitions of expectation and L^ν -spaces gives a sequence of equivalences

$$\mathbb{E}[|\zeta(Z)|^\nu] < \infty \iff \int_{\Omega} |\zeta(z)|^\nu d\mu < \infty \iff \zeta(z) \in L^\nu. \quad (\text{C.41})$$

Lemma C.8. *For a univariate time series $\{Y_t\}_{t \in \mathbb{Z}^+}$ with $\mathbf{Y}_{h:t}$ and $\mathbf{Y}_{\bar{m}:t}$ as defined in definition 2.6, and with m bivariate functions $\zeta_h : \mathbb{R}^2 \rightarrow \mathbb{R}^1$*

$$\begin{aligned} & \text{If } \mathbb{E}[|\zeta_h(\mathbf{Y}_{h:t})|^\nu] < \infty \text{ for } h = 1, \dots, m, \text{ then} \\ & (\mathbb{E}[\sum_{h=1}^m a_h \zeta_h(\mathbf{Y}_{h:t})]^\nu)^{1/\nu} \leq \sum_{h=1}^m |a_h| (\mathbb{E}[|\zeta_h(\mathbf{Y}_{h:t})|^\nu])^{1/\nu} < \infty. \end{aligned}$$

Proof. From eq. (C.41) it follows that $\mathbb{E}[|\zeta_h(\mathbf{Y}_{h:t})|^\nu] < \infty$ implies $\zeta_h(\mathbf{y}_h) \in L_h^\nu$ for $h = 1, \dots, m$. With $\tilde{\zeta}_h(\mathbf{y}_{\bar{m}})$ the corresponding trivial extensions to $(m+1)$ -variate functions, it follows from eq. (2.16) that $\tilde{\zeta}_h(\mathbf{y}_{\bar{m}}) \in L_{\bar{m}}^\nu$ for $h = 1, \dots, m$. From the vector space property of L^ν -spaces it follows that $\sum_{h=1}^m a_h \zeta_h(\mathbf{Y}_{h:t}) \in L_{\bar{m}}^\nu$, and Minkowski's inequality then gives the desired result. \square

C.6 The one-parameter local Gaussian approximation

The input parameter p in the local Gaussian spectral density $f_{v|p}(\omega)$ signifies whether a *five-parameter* or a *one-parameter* local Gaussian approximation has been used in the *local neighbourhood approach* inherited from Hjort and Jones (1996), see section 2 for details. In particular, p is either 5 or 1, and the two different approximation strategies (to be used in the normalised

situation with standard normal marginals) will henceforth be denoted $\psi_5(\mathbf{w})$ or $\psi_1(\mathbf{w})$, where

$$\psi_5(\mathbf{w}; \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) := \frac{1}{2\pi \cdot \sigma_1 \sigma_2 \sqrt{1-\rho^2}} \exp \left\{ -\frac{\sigma_1^2(w_1-\mu_1)^2 - 2\sigma_1\sigma_2\rho(w_1-\mu_1)(w_2-\mu_2) + \sigma_2^2(w_2-\mu_2)^2}{2\sigma_1^2\sigma_2^2(1-\rho^2)} \right\}, \quad (\text{C.42a})$$

$$\psi_1(\mathbf{w}; \rho) := \frac{1}{2\pi \cdot \sqrt{1-\rho^2}} \exp \left\{ -\frac{w_1^2 - 2\rho w_1 w_2 + w_2^2}{2(1-\rho^2)} \right\}. \quad (\text{C.42b})$$

As noted in section 2.1.3, the simplified approach based on $\psi_1(\mathbf{w})$ has been used with good results with regard to density estimation, see Otneim and Tjøstheim (2016, 2017), which motivated that this approach also should be tried in the present paper. However, as was seen in fig. 14, page 33, the local Gaussian spectral densities based on $\rho_{v_{11}}$ gave more or less inconclusive results whereas those based on $\rho_{v_{15}}$ clearly indicated the presence of non-linear traits in the time series under investigation.

It is thus of interest to investigate closer the one-parameter local Gaussian approximation $\psi_1(\mathbf{w})$, i.e. how appropriate will it in general be to use the estimated local correlations obtained from $\psi_1(\mathbf{w})$. It is e.g. evident from eq. (C.42) that $\psi_1(\mathbf{w})$ lacks the flexibility of $\psi_5(\mathbf{w})$, but that is obviously not a detrimental problem with regard to density estimation, as was seen in (Otneim and Tjøstheim, 2016) – where a new density estimation method based on $\psi_1(\mathbf{w})$ in many cases turned out to be just as good or even better than already existing methods. However, the lack of flexibility seems to be a problem when, for a given point \mathbf{v} , the target of interest moves from the estimated density $\hat{\psi}_1(\mathbf{v})$ to the estimated parameter $\hat{\rho}_{v_{11}}$.

The analysis related to $\psi_1(\mathbf{w})$ is straightforward to do, since it in essence only requires that some integrals containing the factor $K_b(\mathbf{w} - \mathbf{v})$ must be reexpressed by the help of substitutions and second order Taylor expansions. The bivariate nature of the problem does however make these expressions a bit cumbersome to work with, so an analogous univariate situation will thus be used instead in order to illustrate the differences between the one- and multi-parameter approach. The idea is that the approach from Tjøstheim and Hufthammer (2013) can be applied on the density estimation method from Hjort and Jones (1996) in a univariate setting, i.e. the target density $g(w)$ will be a univariate distribution whereas the approximating local density will be univariate Gaussian. For simplicity of the analysis, it will henceforth be implicitly assumed that the required regularity assumptions from Tjøstheim and Hufthammer (2013) are satisfied.

Remark C.4. The bivariate $\psi_1(\mathbf{w})$ are used in a situation where an initial step first normalises the marginals of the target density, and then the procedure of density estimation and extraction of the corresponding correlation are performed – which in essence implies that it is the properties of the copula-structure of the target density that are investigated. A similar ‘normalisation of the marginals’ for the univariate framework would however be nonsensical, since that would completely remove all the available information from the case under investigation.

C.6.1 A simplified univariate case – ‘local Gaussian standard deviation’

The problems of interest for the bivariate case ψ_1 and the *local Gaussian correlation* ρ_1 , can be identified from an inspection of the following simplified univariate situation, where the concept of a ‘local Gaussian standard deviation’ can be defined in a completely analogous manner to the one used for the definition of the *local Gaussian correlation*. In particular: For a given

univariate density $g(w)$ and a given point v , first use the univariate Gaussian distribution $\phi(w)$ as the approximating function in the density estimation approach of Hjort and Jones (1996), then define the *local Gaussian standard deviation* σ_v by extracting the standard deviation from the approximating Gaussian distribution.

The univariate analogue of eq. (C.42) is taken to be

$$\phi_2(w; \mu, \sigma) := \frac{1}{\sqrt{2\pi}\cdot\sigma} \exp\left\{-\frac{(w-\mu)^2}{2\sigma^2}\right\}, \quad (\text{C.43a})$$

$$\phi_1(w; \sigma) := \frac{1}{\sqrt{2\pi}\cdot\sigma} \exp\left\{-\frac{w^2}{2\sigma^2}\right\}, \quad (\text{C.43b})$$

where $\phi_1(w; \sigma)$, together with its score function $u_1(w; \sigma) = \frac{d}{d\sigma} \log \phi_1(w; \sigma) = (w^2 - \sigma^2)/\sigma^3$, is the target of interest for the present investigation. It is clear from the rigidity of the one-parameter alternative $\phi_1(w; \sigma)$ that it has some limitations with regard to which values it can take, and it is easy to see that $\phi_1(w; \sigma) \leq \phi_1(w; |w|)$ when $w \neq 0$.

Ignoring for the time being that the inflexibility of $\phi_1(w; \sigma)$ renders it a rather undesirable candidate for density estimation, the univariate version of the procedure explained in eqs. (2.3) and (2.4) can be used for $i = 1, 2$ to minimise the $b \rightarrow 0^+$ limit of the penalty functions

$$q_{v|b|i} = \int K_b(w - v) [\phi_i(w; \sigma) - g(w) \cdot \log \phi_i(w; \sigma)] dw, \quad (\text{C.44})$$

where $K_b(w - v) := \frac{1}{b} \cdot K\left(\frac{w-v}{b}\right)$, with K the standard normal kernel and b the bandwidth. Henceforth focusing on the case $i = 1$, it follows that the value $\sigma_{v|b|1}$ that minimises $q_{v|b|1}$ will satisfy $\frac{d}{d\sigma} q_{v|b|1} = 0$, i.e. the following equation should be satisfied,

$$\int K_b(w - v) u_1(w; \sigma_{v|b|1}) [\phi_1(w; \sigma_{v|b|1}) - g(w)] dw = 0. \quad (\text{C.45})$$

Under the additional assumption that $g(w)$ can be differentiated twice at v , and with $h(w; \sigma)$ defined as $u_1(w; \sigma) [\phi_1(w; \sigma) - g(w)]$ in order to compactify the expressions encountered later on, it follows that the substitution $z = (w - v)/b$ followed by a second order Taylor expansion of $h(w; \sigma)$ around v , enables eq. (C.45) to be restated as

$$h(v; \sigma_{v|b|1}) + b^2 \cdot \frac{1}{2} h^{(2)}(v; \sigma_{v|b|1}) + b^2 \cdot \int R(v + bz; b, \sigma_{v|b|1}) z^2 K(z) dz = 0, \quad (\text{C.46})$$

where R is the remainder term of the Taylor expansion.

It follows from eq. (C.46) that the $b \rightarrow 0^+$ limit of $h(v; \sigma_{v|b|1})$ must be zero. This requirement is satisfied when either $\phi_1(v; \sigma_{v|b|1}) = g(v)$ (preferred case) or $u_1(v; \sigma_{v|b|1}) = 0$ (problematic case). Note in particular that $u_1(v; \sigma_{v|b|1})$ must go to zero in the limit if $g(v) > \phi_1(v; |v|)$, that eq. (C.46) implies that $u_1(v; \sigma_{v|b|1})$ for such a case satisfies

$$u_1(v; \sigma_{v|b|1}) = \frac{b^2 \cdot \left\{ \frac{1}{2} h^{(2)}(v; \sigma_{v|b|1}) + \int R(v + bz; b, \sigma_{v|b|1}) z^2 K(z) dz \right\}}{(g(v) - \phi_1(v; \sigma_{v|b|1}))}, \quad (\text{C.47})$$

and that the $b \rightarrow 0^+$ limit of $\sigma_{v|b|1}$ in this case must be $\sigma_{v|1} = |v|$.

The asymptotic arguments given in Tjøstheim and Hufthammer (2013) can be applied in the present univariate situation too, and it follows that the $n \rightarrow \infty$ limit of the matrices from the Klimko-Nelson approach (see theorem B.1) are given as the following scalar expressions,

$$V_{v|b|1} = b \cdot u_1^2(v; \sigma_{v|b|1}) + b^3 \cdot T_V(v; b, \sigma_{v|b|1}), \quad (\text{C.48a})$$

$$W_{v|b|1} = b \cdot u_1^2(v; \sigma_{v|b|1}) + b^3 \cdot T_W(v; b, \sigma_{v|b|1}), \quad (\text{C.48b})$$

where $T_V(v; b, \sigma_{v|b|1})$ and $T_W(v; b, \sigma_{v|b|1})$ represent the higher order terms and remainder terms of the Taylor expansions used in these cases. The $b \rightarrow 0^+$ and $n \rightarrow \infty$ limit of the asymptotic variance of $\sqrt{n}(\hat{\sigma}_{v|b|1} - \sigma_{v|1})$ is thus based on the $b \rightarrow 0^+$ limit of

$$V_{v|b|1}^{-1} W_{v|b|1} V_{v|b|1}^{-1} = \frac{b \cdot u_1^2(v; \sigma_{v|1}) + b^3 \cdot T_W(v; b, \sigma_{v|b|1})}{[b \cdot u_1^2(v; \sigma_{v|1}) + b^3 \cdot T_V(v; b, \sigma_{v|b|1})]^2}, \quad (\text{C.49})$$

from which it follows that the asymptotic variance of $\sqrt{nb}(\hat{\sigma}_{v|b|1} - \sigma_{v|1})$ is $1/u_1^2(v; \sigma_{v|1})$ when $u_1(v; \sigma_{v|1}) \neq 0$ in the limit. However, as explained above, all cases where $g(v)$ is larger than $\phi_1(v; |v|)$ will give a situation where the limit of $u_1(v; \sigma_{v|1})$ is zero – and it is then clear from the expression for $u_1(v; \sigma_{v|1})$ in eq. (C.47) that the asymptotic result in these cases requires a scaling with $\sqrt{nb^3}$ instead of \sqrt{nb} , and that the asymptotic variance in this case will be the $b \rightarrow 0^+$ limit of $T_W(v; b, \sigma_{v|b|1})/T_V^2(v; b, \sigma_{v|b|1})$.

Remark C.5. The one-parameter approximation $\phi_1(w; \sigma)$ will return the value $\phi_1(v; |v|)$ for all univariate densities $g(w)$ which are twice differentiable at v and satisfies $g(v) > \phi_1(v; |v|)$. This implies that not only does the estimated local standard deviation $\hat{\sigma}_{v|b|1}$ converge slower in these cases, but $\hat{\sigma}_{v|b|1}$ will moreover always converge to the value $\sigma_{v|1} = |v|$. This will happen regardless of the actual value of $g(w)$ at the point v and regardless of how $g(w)$ behaves in a neighbourhood of v , so $\sigma_{v|1}$ does not contain any *local information* about the targeted densities for these cases.

Remark C.6. It is not a problem to use $\phi_1(w; \sigma)$ as a tool for density estimation at v when $g(w)$ satisfies $g(v) < \phi_1(v; |v|)$, as can be seen in fig. 15, where the Beta distribution with parameters $\alpha = 2.19$ and $\beta = 1.50$ has been approximated at the point $v = 0.25$. Both $\phi_1(w; \sigma)$ (blue dashed line) and $\phi_2(w; \sigma)$ (red dotted line) manages to estimate the value $g(0.25) \approx 0.7041$, and are as such equally good as density estimators in this case – however, the corresponding standard deviations, in this case respectively $\sigma_{v|1} \equiv 0.50$ for $\phi_1(w; \sigma)$ and $\sigma_{v|2} \approx 0.29$ for $\phi_2(w; \mu, \sigma)$, can in general be quite different. It might thus be reasonable, even when $\phi_1(w; \sigma)$ can be used to provide a density estimate of $g(w)$ at v , to consider the standard deviation of $\phi_1(w; \sigma)$ to be a somewhat dubious ‘local measure’ of the properties of the target function $g(w)$ in the vicinity of v .

C.6.2 The bivariate case $\psi_1(\mathbf{w}; \rho)$

It is clear from eq. (C.42b) that $\psi_1(\mathbf{0}; \rho)$ never can attain a value below $1/2\pi$, and it is not hard to check that $\psi_1(\mathbf{w}; \rho)$ can attain any positive value when $\mathbf{w} = (w_1, w_2)$ satisfies $|w_1| = |w_2| \neq 0$ (although ρ might need to be very close to -1 or 1 when a high value is desired). For other

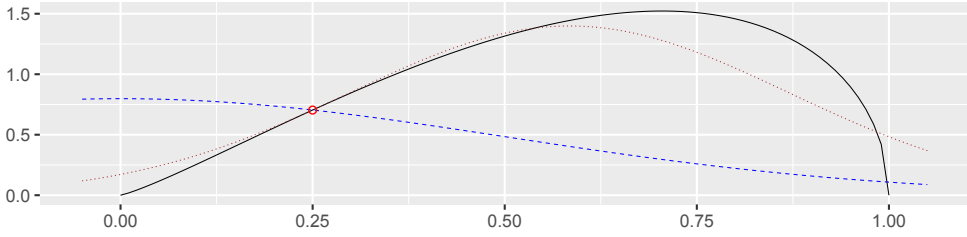


Figure 15: Simplified univariate investigation based on ‘local Gaussian standard deviations’.

points \mathbf{w} than those already mentioned, an inspection of the corresponding score function,

$$u_1(\mathbf{w}; \rho) = \frac{d}{d\rho} \log(\psi_1(\mathbf{w}; \rho)) = \frac{-\rho^3 + \rho^2 w_1 w_2 + \rho(1 - (w_1^2 + w_2^2)) + w_1 w_2}{(1 - \rho^2)^2}, \quad (\text{C.50})$$

together with the requirement that ρ lies in $(-1, 1)$, gives that there is an upper limit $\psi_1(\mathbf{w}; \rho_{\mathbf{w}|1}^\circ)$ for $\psi_1(\mathbf{w}; \rho)$ at this particular point. Note that $\rho_{\mathbf{w}|1}^\circ$ is a solution of $u_1(\mathbf{w}; \rho) = 0$, and these solutions can be explicitly computed by the help of the formula for the roots of the cubic function, but this is not necessary to do for the present discussion.

The arguments used for the one-parameter univariate $\phi_1(w, \sigma)$, see eqs. (C.44) to (C.49), can be extended directly to the one-parameter bivariate $\psi_1(\mathbf{w})$. The observations below follows from those given in remarks C.5 and C.6. Note that the bivariate target density $g(\mathbf{w})$ here corresponds to a copula-model with standard normal margins, and $\mathbf{v} = (v_1, v_2)$ represents the point at which the density estimation and extraction of correlation $\rho_{\mathbf{v}|1}$ is performed.

- (a) If $\mathbf{v} = \mathbf{0}$ and $g(\mathbf{0}) < 1/2\pi$, then $\rho_{\mathbf{v}|1} = 0$ and the estimated density always becomes $\psi_1(\mathbf{0}; 0) = 1/2\pi$. The estimates of $\rho_{\mathbf{v}|1}$ will in this case converge towards $\rho_{\mathbf{v}|1} = 0$ with a slower convergence rate than $\sqrt{nb_1 b_2}$, and $\rho_{\mathbf{v}|1} = 0$ will always occur when $g(\mathbf{0}) < 1/2\pi$, regardless of how $g(\mathbf{w})$ behaves in a neighbourhood of $\mathbf{v} = \mathbf{0}$.
- (b) If \mathbf{v} ‘lies off the diagonals’, i.e. $|v_1| \neq |v_2|$, and $g(\mathbf{v}) > \psi_1(\mathbf{v}; \rho_{\mathbf{v}|1}^\circ)$, then $\rho_{\mathbf{v}|1} = \rho_{\mathbf{v}|1}^\circ$ and the estimated density always becomes $\psi_1(\mathbf{v}; \rho_{\mathbf{v}|1}^\circ)$. The estimates of $\rho_{\mathbf{v}|1}$ will in this case converge towards $\rho_{\mathbf{v}|1}^\circ$ with a slower convergence rate than $\sqrt{nb_1 b_2}$, and $\rho_{\mathbf{v}|1} = \rho_{\mathbf{v}|1}^\circ$ will always occur when $g(\mathbf{v}) > \psi_1(\mathbf{v}; \rho_{\mathbf{v}|1}^\circ)$, regardless of how $g(\mathbf{w})$ behaves in a neighbourhood of \mathbf{v} .
- (c) The one-parameter approximation $\psi_1(\mathbf{w}; \rho)$ will give the correct density estimate of $g(\mathbf{w})$ at \mathbf{v} for other cases than those excluded above, but the corresponding correlation $\rho_{\mathbf{v}|1}$ might differ substantially from the one obtained from the five-parameter approximation.

The case in item (b) can e.g. be observed when $g(\mathbf{w})$ is the probability density function of a Clayton copula with standard normal marginals.

Remark C.7. The preceding discussion assumed that the target density $g(\mathbf{w})$ was known, and this will in general not be the case when a sample is investigated. Small sample variation must also be taken into account when local Gaussian autocorrelations are to be computed for a given sample, and it might be hard to test whether or not there are significant differences between the correlations obtained from the five- and one-parameter approximations. However, clear differences can occur, like seen in fig. 14 (page 33), where the estimates of $f_{\mathbf{v}|5}(\omega)$ and $f_{\mathbf{v}|1}(\omega)$ for the dmbp-case was discussed. It might thus be enlightening to briefly return to the dmbp-

case and present a plot that shows how the estimated local Gaussian autocorrelations based on the one- and five-parameter approaches looks like, see fig. 16. The point 90% : 90% is the same that was used in fig. 14, the red dotted lines shows the truncation level $m = 10$ used for the estimated local Gaussian spectra, and all the estimations have as usual been performed based on the bandwidth $\mathbf{b} = (0.5, 0.5)$. As can be seen, the estimates based on the $\psi_1(\mathbf{w})$ -approximation tend to have the same sign as those from the $\psi_5(\mathbf{w})$ -approximation, but the values are in general much closer to 0.

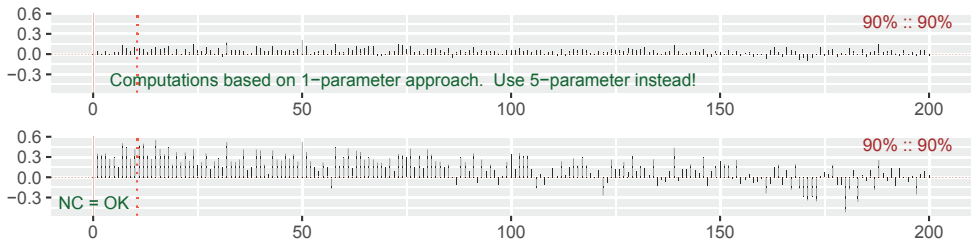


Figure 16: A comparison of estimated local Gaussian autocorrelations from the dmbp-data: $\hat{\rho}_{\psi_1}(h)$ from the $\psi_1(\mathbf{w})$ -approximation (top) versus $\hat{\rho}_{\psi_5}(h)$ from the $\psi_5(\mathbf{w})$ -approximation (bottom).

Remark C.8. Keeping in mind that the one- and five-parameter local Gaussian approximations both should return approximately the same estimated correlation-values when the density $g_h(\mathbf{y}_h)$ of (Y_{t+h}, Y_t) either is Gaussian or independent, it seems plausible that a visual comparison like the one in fig. 16 might be used to provide some insight into the long range behaviour of the time series under interest. However, as discussed in section 4, it is important to keep in mind that there are uncertainties related to the level of small sample variation that can occur for different configuration of the input parameters.

Remark C.9. It might be the case that the ‘trumpet shape’ that occurred for some of the estimated pointwise confidence bands in fig. 14 could be related to a situation where the density $g_h(\mathbf{y}_h)$ behaves in such a manner that the one-parameter approximation $\psi_1(\mathbf{w}; \rho)$ runs into problems as a density estimator. This has not been investigated in this paper, but it should be possible to approach this idea by a two step procedure where 1) copula models are fitted to the pairs (Y_{t+h}, Y_t) , and 2) the selected copula-models (with standard normal marginals) are compared against the one-parameter case $\psi_1(\mathbf{w}; \rho)$.

Remark C.10. Figures 14 and 16 are the only plots in this paper that compares the results based on correlations extracted from $\psi_1(\mathbf{w})$ with those based on correlations from $\psi_5(\mathbf{w})$. The scripts stored in the R-package `localgaussSpec` can be used to get hold of additional examples. Use `devtools::install_github("LAJordanger/localgaussSpec")` to install the package, then load the package and use `LG_extract_scripts` to get access to the scripts. The resulting plots can be interactively investigated by means of the `shiny`-application part of the package, read the scripts and the package-documentation for further details.

References

- Ahdesmäki, M., Lähdesmäki, H., Pearson, R., Huttunen, H., Yli-Harja, O., 2005. Robust detection of periodic time series measured from biological systems. *BMC bioinformatics* 6 (1), 117.
- Akaike, H., 1966. Note on higher order spectra. *Annals of the Institute of Statistical Mathematics* 18 (1), 123–126.
URL <http://dx.doi.org/10.1007/BF02869523>
- 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 (2), 352–380.
URL <http://dx.doi.org/10.1111/jtsa.12183>
- Berentsen, G. D., Kleppe, T. S., Tjøstheim, D. B., Feb. 2014a. Introducing `localgauss`, an R Package for Estimating and Visualizing Local Gaussian Correlation. *j-J-STAT-SOFT* 56 (12).
URL <http://www.jstatsoft.org/v56/i12>
- Berentsen, G. D., Tjøstheim, D., 2014. Recognizing and visualizing departures from independence in bivariate data using local Gaussian correlation. *Statistics and Computing* 24 (5), 785–801.
URL <http://dx.doi.org/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>
- Billingsley, P., 2012. *Probability and Measure, Anniversary Edition*. Wiley.
- Bollerslev, T., Ghysels, E., 1996. Periodic Autoregressive Conditional Heteroscedasticity. *Journal of Business & Economic Statistics* 14 (2), 139–151.
URL <http://amstat.tandfonline.com/doi/abs/10.1080/07350015.1996.10524640>
- Brillinger, D. R., 1965. An Introduction to Polyspectra. *The Annals of Mathematical Statistics* 36 (5), 1351–1374.
URL <http://www.jstor.org/stable/2238424>
- 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 (465-476), 24J.
URL <http://www3.stat.sinica.edu.tw/statistica/j1n2/j1n23/..j1n210\j1n210.htm>

- 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 (3), 295–310.
URL <http://www.sciencedirect.com/science/article/pii/S0167947399000146>
- Burman, P., Chow, E., Nolan, D., 1994. A Cross-Validatory Method for Dependent Data. *Biometrika* 81 (2), 351–358.
URL <http://www.jstor.org/stable/2336965>
- Carcea, M., Serfling, R., 2015. A Gini Autocovariance Function for Time Series Modelling. *Journal of Time Series Analysis* 36 (6), 817–838, 10.1111/jtsa.12130.
URL <http://dx.doi.org/10.1111/jtsa.12130>
- Chang, W., Cheng, J., Allaire, J., Xie, Y., McPherson, J., 2017. shiny: Web Application Framework for R. R package version 1.0.3.
URL <https://CRAN.R-project.org/package=shiny>
- Davydov, Y. A., 1968. Convergence of Distributions Generated by Stationary Stochastic Processes. *Theory of Probability and Application* 13 (4), 691–696.
URL <http://dx.doi.org/10.1137/1113086>
- Dette, H., Hallin, M., Kley, T., Volgushev, S., 05 2015. Of copulas, quantiles, ranks and spectra: An L_1 -approach to spectral analysis. *Bernoulli* 21 (2), 781–831.
URL <http://dx.doi.org/10.3150/13-BEJ587>
- 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 (1), 83–106.
URL <http://www.sciencedirect.com/science/article/pii/S092753989390006D>
- Fan, J., Yao, Q., 2003. *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer.
- Ghalanos, A., 2015a. Introduction to the rugarch package (Version 1.3-1).
URL https://CRAN.R-project.org/web/packages/rugarch/vignettes/Introduction_to_the_rugarch_package.pdf
- Ghalanos, A., 2015b. rugarch: Univariate GARCH models. R package version 1.3-6.
URL <https://cran.r-project.org/package=rugarch>
- Hagemann, A., November 2011. Robust Spectral Analysis.
URL <https://ssrn.com/abstract=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 (1), 251 – 270.
URL <http://www.sciencedirect.com/science/article/pii/S0304407616300458>

- Hjort, N. L., Jones, M. C., 08 1996. Locally parametric nonparametric density estimation. *Ann. Statist.* 24 (4), 1619–1647.
URL <http://dx.doi.org/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 (448), 1201–1220.
URL <http://tandfonline.com/doi/abs/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 (3), 557–574.
URL <http://onlinelibrary.wiley.com/doi/10.1111/1467-9868.00250/abstract>
- Hong, Y., Tu, J., Zhou, G., 2007. Asymmetries in Stock Returns: Statistical Tests and Economic Evaluation. *The Review of Financial Studies* 20 (5), 1547–1581.
URL <http://www.jstor.org/stable/4494812>
- Horn, R. A., Johnson, C. R., 2012. *Matrix Analysis*, 2nd Edition. Cambridge University Press, New York, NY, USA.
- Klimko, L. A., Nelson, P. I., 05 1978. On Conditional Least Squares Estimation for Stochastic Processes. *Ann. Statist.* 6 (3), 629–642.
URL <http://dx.doi.org/10.1214/aos/1176344207>
- Koenker, R., 2005. *Quantile Regression*. Vol. 38 of *Econometric Society Monographs*. Cambridge University Press.
- Koenker, R., Bassett Jr, G., 1978. Regression Quantiles. *Econometrica* 46 (1), 33–50.
URL <http://www.jstor.org/stable/1913643>
- Lacal, V., Tjøstheim, D., 2017. Local Gaussian Autocorrelation and Tests for Serial Independence. *Journal of Time Series Analysis* 38 (1), 51–71, 10.1111/jtsa.12195.
URL <http://dx.doi.org/10.1111/jtsa.12195>
- 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 (3), 292–321.
URL <http://www.sciencedirect.com/science/article/pii/S1572312706000505>
- 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>
- Li, T.-H., 2008. Laplace Periodogram for Time Series Analysis. *Journal of the American Statistical Association* 103 (482), 757–768.
URL <http://dx.doi.org/10.1198/016214508000000265>

- Li, T.-H., May 2010a. A Nonlinear Method for Robust Spectral Analysis. *IEEE Transactions on Signal Processing* 58 (5), 2466–2474.
URL <http://ieeexplore.ieee.org/abstract/document/5406102/>
- Li, T.-H., Aug 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 (7), 2133 – 2140.
URL <http://www.sciencedirect.com/science/article/pii/S0165168410000137>
- Li, T.-H., March 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/>
- Li, T.-H., 2012b. On robust spectral analysis by least absolute deviations. *Journal of Time Series Analysis* 33 (2), 298–303.
URL <http://dx.doi.org/10.1111/j.1467-9892.2011.00760.x>
- Li, T.-H., 2012c. Quantile Periodograms. *Journal of the American Statistical Association* 107 (498), 765–776.
URL <http://dx.doi.org/10.1080/01621459.2012.682815>
- Li, T.-H., 2014. Quantile Periodogram and Time-Dependent Variance. *Journal of Time Series Analysis* 35 (4), 322–340.
URL <http://dx.doi.org/10.1111/jtsa.12065>
- Linton, O., Whang, Y.-J., 2007. The quantilogram: With an application to evaluating directional predictability. *Journal of Econometrics* 141 (1), 250 – 282, semiparametric methods in econometrics.
URL <http://www.sciencedirect.com/science/article/pii/S0304407607000152>
- 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 (02), 258–289.
URL http://EconPapers.repec.org/RePEc:cup:etheor:v:11:y:1995:i:02:p:258-289_00
- Nelsen, R. B., 2006. *An Introduction to Copulas -*, 2nd Edition. Springer, Berlin, Heidelberg.
- Nordman, D. J., Lahiri, S. N., 05 2014. Convergence rates of empirical block length selectors for block bootstrap. *Bernoulli* 20 (2), 958–978.
URL <http://dx.doi.org/10.3150/13-BEJ511>
- Otneim, H., Tjøstheim, D., Oct 2016. The locally Gaussian density estimator for multivariate data. *Statistics and Computing*, 1–22.
URL <https://doi.org/10.1007/s11222-016-9706-6>

- Otneim, H., Tjøstheim, D., 2017. Conditional density estimation using the local Gaussian correlation. *Statistics and Computing*, 1–19.
URL <http://dx.doi.org/10.1007/s11222-017-9732-z>
- 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 (4), 372–375.
URL <http://dx.doi.org/10.1080/07474930802459016>
- 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 (2), 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 (1), 53–70.
URL <http://dx.doi.org/10.1081/ETC-120028836>
- Racine, J., 2000. Consistent cross-validators model-selection for dependent data: *h_v*-block cross-validation. *Journal of Econometrics* 99 (1), 39–61.
URL <http://www.sciencedirect.com/science/article/pii/S0304407600000300>
- Shao, J., 1993. Linear Model Selection by Cross-validation. *Journal of the American statistical Association* 88 (422), 486–494.
URL <http://www.tandfonline.com/doi/abs/10.1080/01621459.1993.10476299>
- 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.
- Taniguchi, M., Kakizawa, Y., 2000. *Asymptotic Theory of Statistical Inference for Time Series*. Springer.
- Teräsvirta, T., Tjøstheim, D., Granger, C. W., et al., 2010. *Modelling nonlinear economic time series*. OUP Catalogue.
- Tjøstheim, D., Hufthammer, K. O., 2013. Local Gaussian correlation: A new measure of dependence. *Journal of Econometrics* 172 (1), 33 – 48.
URL <http://www.sciencedirect.com/science/article/pii/S0304407612001741>
- Tong, H., 1990. *Non-linear time series: a dynamical system approach*. Oxford University Press.
- 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.
- Van Hecke, R., Volgushev, S., Dette, H., 2017. Fourier analysis of serial dependence measures. arXiv preprint arXiv:1703.04320.
URL <https://arxiv.org/abs/1703.04320>

- Volkonskii, V. A., Rozanov, Y. A., 1959. Some Limit Theorems for Random Functions. I. Theory of Probability and Application 4 (2), 178–197.
URL <http://epubs.siam.org/doi/10.1137/1104015>
- Wang, X., Hong, Y., 2017. Characteristic function based testing for conditional independence: A nonparametric regression approach. Econometric Theory, 1–35.
URL <https://doi.org/10.1017/S026646661700010X>
- Xie, Y., 2015. Dynamic Documents with R and knitr, 2nd Edition. Chapman and Hall/CRC, Boca Raton, Florida, ISBN 978-1498716963.
URL <http://yihui.name/knitr/>
- Xie, Y., 2016. knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.15.1.
URL <http://yihui.name/knitr/>
- Zhou, Z., 2012. Measuring nonlinear dependence in time-series, a distance correlation approach. Journal of Time Series Analysis 33 (3), 438–457.
URL <http://dx.doi.org/10.1111/j.1467-9892.2011.00780.x>