Nonlinear Spectrum Analysis based on the Local Gaussian Correlation and Model Selection for Copulas

Lars Arne Jordanger

Thesis for the degree of philosophiae doctor (PhD)
at the University of Bergen

2017

Date of defence: November 20th 2017

Preface

The present work has been carried out during my employment as PhD-student at the Department of Mathematics at the University of Bergen, Norway, lasting from August 2013 until August 2017.

This thesis consists of two parts. The first part gives a brief introduction to the main ideas and topics of the thesis, whereas the second part consists of the following papers:

Paper 1 Lars Arne Jordanger and Dag Tjøstheim, 'Nonlinear spectral analysis via the local Gaussian correlation'. This paper is available at arXiv.org, https://arxiv.org/abs/1708.02166.

Paper 2 Lars Arne Jordanger and Dag Tjøstheim, 'Nonlinear cross-spectrum analysis via the local Gaussian correlation'. This paper is available at arXiv.org, https://arxiv.org/abs/1708.02495.

Paper 3 Lars Arne Jordanger and Dag Tjøstheim, 'Model selection of copulas: AIC versus a cross validation copula information criterion', *Statistics and Probability Letters*, **92**, 249-255 (2014). Copyright (2014) Elsevier B.V.[1]

[1]The published paper is reprinted with permission from Elsevier B.V., cf. https://www.elsevier.com/__data/assets/pdf_file/0007/55654/AuthorUserRights.pdf

Acknowledgements

First of all, I would like to thank my supervisor, Professor Dag Tjøstheim, for his patience and support during my work on this project. Tjøstheim was also my supervisor for my Master's thesis, so altogether I have had the pleasure and privilege of cooperating with him for the last six years, during which we have had a lot of interesting discussions about the statistical theories that this PhD thesis are based upon. I initially never considered a PhD in statistics to be an option of interest for me, but I changed my mind due to the excellent and enthusiastic supervision I received from professor Tjøstheim during my Master's Degree. Thank you very much for your support during all these years.

I would also like to thank my colleagues and the academic staff at the Department of Mathematics, University of Bergen. It has been a great pleasure, both as a student and as an employee, to be a part of this department.

In addition, I want to express my gratitude towards the open source software community for all the magnificent tools they have created. My work on this project has in particular benefited from the following programs/packages; R, knitr, LaTeX, git, emacs, ess and magit. These programs have made the writing and programming related to this project tremendously easier to manage, and I only regret that I did not start to use git and magit at an earlier stage – since in hindsight I realise that that would have saved me for several weeks worth of mundane and tedious bookkeeping tasks.

Finally, I would like to thank my friends and family for being there, and in particular my dear parents (Gerd and Johannes) for their love and support.

Bergen, August 2017
Lars Arne Jordanger

Abstracts

Paper 1 'Nonlinear spectral analysis via the local Gaussian correlation'
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* $\rho_{\boldsymbol{v}}(h)$ introduced in Tjøstheim and Hufthammer (2013), and these *local measures of dependence* can be used to construct the *local Gaussian spectral density* $f_{\boldsymbol{v}}(\omega)$ that is presented in this paper. A key feature of $f_{\boldsymbol{v}}(\omega)$ is that it coincides with $f(\omega)$ for Gaussian time series, which implies that $f_{\boldsymbol{v}}(\omega)$ 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 $f_{\boldsymbol{v}}(\omega)$ can indicate nonlinear traits, which potentially might discover *local periodic phenomena* that goes undetected in an ordinary spectral analysis.

Paper 2 'Nonlinear cross-spectrum analysis via the local Gaussian correlation'
Spectrum analysis can detect frequency related structures in a time series $\{\boldsymbol{Y}_t\}_{t\in\mathbb{Z}}$, but may in general be an inadequate tool if asymmetries or other nonlinear phenomena are present. This limitation is a consequence of the way the spectrum is based on the second order moments (auto and cross-covariances), and alternative approaches to spectrum analysis have thus been investigated based on other measures of dependence. One such approach was developed for univariate time series in Jordanger and Tjøstheim (2017), where it was seen that a *local Gaussian auto-spectrum* $f_{\boldsymbol{v}}(\omega)$, based on the *local Gaussian autocorrelations* $\rho_{\boldsymbol{v}}(h)$ from Tjøstheim and Hufthammer (2013), could detect local structures in time series that looked like white noise when investigated by the ordinary auto-spectrum $f(\omega)$. The *local Gaussian approach* in this paper is extended to a *local Gaussian cross-spectrum* $f_{k\ell:\boldsymbol{v}}(\omega)$ for multivariate time series. The local cross-spectrum $f_{k\ell:\boldsymbol{v}}(\omega)$ has the desirable property that it coincides with the ordinary cross-spectrum $f_{k\ell}(\omega)$ for Gaussian time series, which implies that $f_{k\ell:\boldsymbol{v}}(\omega)$ can be used to detect non-Gaussian traits in the time series under investigation. In particular: If the ordinary spectrum is flat, then peaks and troughs of *the local Gaussian spectrum* can indicate

nonlinear traits, which potentially might discover *local periodic phenomena* that goes undetected in an ordinary spectral analysis.

Paper 3 'Model selection of copulas: AIC versus a cross validation copula information criterion'
Akaikes Information Criterion (AIC) is frequently employed in the semiparametric setting of selection of copula models, even though as a model selection tool it was developed in a parametric setting. Recently a Copula Information Criterion (CIC) has been especially designed for copula model selection. In this paper we examine the two approaches and present a simulation study where the performance of a cross-validated version of CIC is compared with the AIC criterion. Only minor differences are observed.

Contents

Chapter 1

Introduction

For a time series $\{Y_t\}_{t\in\mathbb{Z}}$, the auto- and cross-spectra can only detect those properties that are captured by the auto- and cross-covariances, and an investigation of features that goes unnoticed by these second order moments will require some other approach.

This thesis presents new tools for the detection and investigation of nonlinear structures, in particular periodicities, in time series, i.e. the *local Gaussian auto- and cross-spectra* (Paper 1 and 2, respectively). These tools are based on a *local Gaussian* investigation of the probability density functions related to the lag h pairs $\left(Y_{t+h}, Y_t\right)$ of the time series under investigation, and a *local Gaussian spectral analysis* of $\{Y_t\}_{t\in\mathbb{Z}}$ can thus be conducted – which potentially might discover periodic phenomena that goes undetected in an ordinary spectral analysis.

The *local Gaussian spectra* might also be of interest with regard to an informal model selection for time series, as they enable a local view of the suitability of different parametric models fitted to a given time series. The model selection aspect is taken up more systematically, but restricted to a copula setting, in the last paper of this thesis (Paper 3), where a simulation study compares different model selection tools for copula-models.

A unifying theoretical framework which many of the concepts of this thesis builds upon will be briefly outlined and discussed in this introduction. It will in particular be seen that all the papers are in some sense related to the Kullback-Leibler information criterion (KLIC), that a bit of copula-theory is needed, and that transformations of the observations plays a central role.

1.1 Theoretical framework: The basic components

1.1.1 Extraction of information from observations

The detection of structure in a given set of n observations, say $\mathcal{X}_n = \{\boldsymbol{x}_i\}_{i=1}^n$ with $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{id})$, requires techniques for the extraction of the inherent information. In some cases it might be sufficient to let the *data speak for themselves*, e.g. non-parametric methods can be applied in order to get a density estimate. In other cases it might be better to look for a parametric or semi-parametric model, in which case it will be necessary

to select a collection of models $\mathcal{F} \coloneqq \{F_i\}_{i=1}^{s}$ (with probably density functions f_i and general parameters $\boldsymbol{\alpha}_i$), and then use an information criterion to select the model $F_i^\star$ (with optimal parameters $\boldsymbol{\alpha}_i^\star$) that gives the *best match* with the available observations $\mathcal{X}_n$.

The concept of *best match* can be a tricky one, i.e. since the different selection methods might rest upon assumptions that could be hard to verify (based on the available data). A brief reminder of the underlying theory can thus be enlightening.

1.1.2 The Kullback-Leibler divergence and information criteria

Many information criteria, like the *Akaike Information Criterion* (AIC) and the *Takeuchi Information Criterion* (TIC), see respectively Akaike (1974); Takeuchi (1976), are related to different estimation regimes of the *Kullback-Leibler Information Criterion* (KLIC), see Kullback and Leibler (1951). In particular, assuming that the sample $\mathcal{X}_n$ originates from a *known model* F° (with probability density function f°), then it is possible to compute the following entity for each model $F_i \in \mathcal{F}$,

$$\text{KLIC}(F^\circ, F_i) \coloneqq \text{E}_{f^\circ}\left[\log \frac{f^\circ}{f_i}\right] = \text{E}_{f^\circ}[\log f^\circ] - \text{E}_{f^\circ}[\log f_i]\,. \tag{1.1}$$

It follows from the Jensen inequality that $\text{KLIC}(F^\circ, F_i)$ always will be non-negative, and that it is equal to zero if and only if $f_i = f^\circ$ almost everywhere – and this implies that $\text{KLIC}(F^\circ, F_i)$ can be used to find the optimal parameter-configuration $\boldsymbol{\alpha}_i^\star$ that makes $F_i^\star$ the *best match* to F°. Equation (1.1) can then be used to select the *overall best model* $F^\star$ from the collection of *best matches* $\{F_i^\star\}_{i=1}^{s}$. It is important to note that it might happen that none of the models in $\mathcal{F}$ gives a decent approximation of F° – and in that case it will be a bit of a misnomer to refer to $F^\star$ as the *'overall best model'*.

Since the term $\text{E}_{f^\circ}[\log f^\circ]$ is present in eq. (1.1) for all the models, it is clear that a minimisation of $\text{KLIC}(F^\circ, F_i)$ corresponds to a maximisation of $\text{E}_{f^\circ}[\log f_i]$, and many information criteria are based on the strategy that this latter term are estimated from the observations $\mathcal{X}_n$. The actual estimation will be influenced by the assumptions made upon the observations $\mathcal{X}_n$ and the models $\mathcal{F}$, which leads to different information criteria, like the before mentioned AIC and TIC. These model selection strategies will be subject to the effect of small-sample variation, and additional complications might also occur if they are applied *outside* of the realm in which they originally were constructed.

1.1.3 The copula and the selection of copula models

Sklar's theorem, see Sklar (1959), states that a multivariate model $F_i(x_1, \ldots, x_d)$ can be written as

$$F_i(x_1, \ldots, x_d) = C_i(F_{i1}(x_1), \ldots, F_{id}(x_d)) \tag{1.2}$$

where $\left\{F_{ij}\left(x_j\right)\right\}_{j=1}^{d}$ are the d marginal cumulative distributions, and where $C_i(u_1, \ldots, u_d)$ is a d-variate copula that decodes all the interdependence structure of $F_i(x_1, \ldots, x_d)$.

Equation (1.2) enables a decoupling of the parameters of F_i into those that occurs in the marginal distributions F_{ij}, and those that occur in the copula C_i. The parameters of the marginals are often considered as nuisance parameters if the aim of the investigation is to find the copula-model that best describes the internal dependency structure of a given sample $\mathcal{X}_n$. It is then natural to use a semi-parametric two-step approach, where the first step uses the $\tilde{F}_i$ (the $\frac{n}{n+1}$-scaled marginal empirical cumulative density functions) to *transform from observations into pseudo-uniform observations*, and the second step is to find the copula model $C^\star$ (from the collection $\mathcal{C} = \{C_i\}_{i=1}^s$) that best matches the dependency structure of the pseudo-uniform observations.

Grønneberg and Hjort (2008) showed that the transformation to pseudo-uniform observations had a non-negligible impact upon the asymptotic behaviour of the estimator, i.e. that it formally was not quite correct to use a parametric method like AIC in this semi-parametric setting. An investigation of the $\mathrm{E}_{f^\circ}[\log f_i]$ term from eq. (1.1), that properly took into account the effect of the pseudo-transformation of the observed data $\mathcal{X}_n$, lead to different information criteria for copula-selection, see Grønneberg and Hjort (2008, 2014). The *cross-validation Copula Information Criterion* (xv-CIC) from (Grønneberg and Hjort, 2014) is the topic for Paper 3 in this thesis.

The copula also plays an important indirect role in Papers 1 and 2, with the introduction of pairs of transformed pseudo-normal variables, see eq. (1.4) in section 1.1.5 for further details.

1.1.4 Local parametric models and the local Gaussian correlation $\rho_{\boldsymbol{v}}$

It might be a tall order to find a *reasonable* collection of parametric models $\mathcal{F} = \{F_i\}_{i=1}^s$ when a data set $\mathcal{X}_n$ is to be investigated, in particular if the desired application is to estimate f° at a given point $\boldsymbol{v} \in \mathbb{R}^d$. The core of the problem is that even though eq. (1.1) is used to select an optimal model $F^\star$ from $\mathcal{F}$, it might still be the case that $f^\star(\boldsymbol{v})$ fails to give a decent estimate of $f^\circ(\boldsymbol{v})$. The reason for this is that $F^\star$ has been selected based on the property that its density $f^\star$ minimises the expectation of $\log(f^\circ/f_i)$, and this does not guarantee a good match between f° and $f^\star$ at a specified point $\boldsymbol{v}$.

A solution to this problem was introduced in Hjort and Jones (1996), where a *locally parametric nonparametric density estimation* approach was presented. In this approach, for a given parametric model $F(\boldsymbol{x};\boldsymbol{\theta})$, the aim is to find a solution $F(\boldsymbol{x};\boldsymbol{\theta}_{\boldsymbol{v}})$ that minimises the $\boldsymbol{b} \to \boldsymbol{0}^+$ limit of the following locally weighted version of KLIC, where $K_{\boldsymbol{b}}(\boldsymbol{x}-\boldsymbol{v})$ is a kernel function and $\boldsymbol{b}$ is the bandwidth,

$$\int K_{\boldsymbol{b}}(\boldsymbol{x}-\boldsymbol{v})\left[f^\circ(\boldsymbol{x})\log\tfrac{f^\circ(\boldsymbol{x})}{f(\boldsymbol{x};\boldsymbol{\theta})} - \{f^\circ(\boldsymbol{x}) - f(\boldsymbol{x};\boldsymbol{\theta})\}\right]\mathrm{d}\boldsymbol{x}. \tag{1.3}$$

Given that the model $F(\boldsymbol{x};\boldsymbol{\theta})$ is *sufficiently flexible*, this approach has a better chance of ensuring a result that satisfies the requirement $f^\circ(\boldsymbol{v}) = f(\boldsymbol{v};\boldsymbol{\theta}_{\boldsymbol{v}})$.

The aim of (Hjort and Jones, 1996) was the estimation of the density $f^\circ(\boldsymbol{v})$, but it is also possible to change the focus to the estimated parameter vector $\boldsymbol{\theta}_{\boldsymbol{v}}$. This approach

was pursued for the bivariate case in Tjøstheim and Hufthammer (2013), where the specification that $F(\boldsymbol{x};\boldsymbol{\theta})$ should be a Gaussian distribution enabled the extraction of the *local Gaussian correlation* ρ_v from $\boldsymbol{\theta}_v$. The ρ_v obtained in this manner provides a *local measure of dependence*, with the desirable properties that it *(a)* will return the ordinary (global) correlation if the targeted function F° itself is Gaussian, and that it *(b)* can detect nonlinear dependencies between uncorrelated random variables.

The *local Gaussian approach* from (Tjøstheim and Hufthammer, 2013) has been investigated in different settings, some focus (like (Hjort and Jones, 1996)) on density estimation, see Otneim and Tjøstheim (2016, 2017), whereas other investigate properties of the *local Gaussian correlation* as a *local measure of dependence*, see Berentsen et al. (2017, 2014a); Berentsen and Tjøstheim (2014); Berentsen et al. (2014b); Lacal and Tjøstheim (2016, 2017); Støve and Tjøstheim (2014); Støve et al. (2014).

The idea that the *local Gaussian correlation* should be applied to time series was an integral part of (Tjøstheim and Hufthammer, 2013), i.e. the theory was developed for the case with lag h pairs $\left(Y_{t+h}, Y_t\right)$ from a univariate time series $\{Y_t\}_{t\in\mathbb{Z}}$. This case will in Paper 1 be referred to as $\rho_v(h)$, the *local Gaussian autocorrelation at lag* h, and the basic idea is to use the collection $\{\rho_v(h)\}_{h\in\mathbb{Z}}$ to define the *local Gaussian auto-spectral density* for a univariate time series. The extension to multivariate time series, and the concept of *local Gaussian cross-spectral density*, is given in Paper 2.

1.1.5 Estimation of the local Gaussian correlation ρ_v

The R-package `localgauss`, see (Berentsen et al., 2014a), can compute estimates of the local Gaussian correlation for bivariate samples. Figure 1.1 shows two examples, both with samples of size $n = 500$. The sample in the left panel is from a bivariate normal distribution with correlation 0.5 and standard normal margins. In this case the local Gaussian correlation coincides with the ordinary (global) correlation, and this agrees (taking small-sample variation into account) with the observed estimates of the local Gaussian correlation. The case investigated in the right panel is from a bivariate random variable (X_1, X_2), where $X_2 = X_1^2 + \epsilon$ with $X_1 \sim \mathrm{N}(0,1)$ and $\epsilon \sim \mathrm{N}\left(0, \frac{1}{4}\right)$. The random variables X_1 and X_2 are uncorrelated, but obviously dependent, as can be see in the right panel of fig. 1.1. For this sample the estimated correlation is -0.015, which does not reveal anything about the internal dependency structure between X_1 and X_2. On the other hand, the estimated local Gaussian correlations clearly detect significant local features of the underlying distribution.

The estimated local Gaussian correlations in fig. 1.1 are based on the *original observations* $\{(x_{1i}, x_{2i})\}_{i=1}^n$, but it might in general be necessary to transform to *pseudo-normalised observations* in order to avoid numerical problems during the estimation process. The *pseudo-normalisation procedure* used in Paper 1 and 2 is a two step process, where *(I)* the first step is to use the $\tilde{F}_i$ (the $\frac{n}{n+1}$-scaled empirical marginal cumulative density functions) to replace the observations with *pseudo-uniform observations*, and *(II)* the second step is to use Φ^{-1} (the inverse of the cumulative density function for the univariate standard normal distribution) in order to get the *pseudo-normalised*

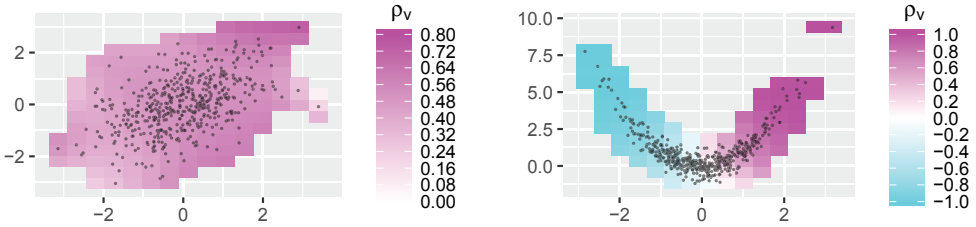

Figure 1.1: Examples of the local Gaussian correlation ρ_v.

observations, i.e.

$$\{(x_{1i}, x_{2i})\}_{i=1}^{n} \overset{I}{\rightsquigarrow} \left\{\left(\tilde{F}_1(x_{1i}), \tilde{F}_2(x_{2i})\right)\right\}_{i=1}^{n} \overset{II}{\rightsquigarrow} \left\{\left(\Phi^{-1}(\tilde{F}_1(x_{1i})), \Phi^{-1}(\tilde{F}_2(x_{2i}))\right)\right\}_{i=1}^{n}. \quad (1.4)$$

This pseudo-normalisation procedure in essence makes it possible to ignore the effect of the marginal distributions in the subsequent analysis, and puts the emphasis on the copula-component that describes the interdependence of the variables. It is important to note that this transformation does not destroy the desirable property that the local Gaussian correlation of a *Gaussian distribution* will be equal to the correlation of *that* distribution – which is pivotal for the interpretations that can be made later on in the discussion of *the local Gaussian spectral densities*.

1.1.6 The local Gaussian correlation ρ_v versus quantile based local measures

There are several strategies available that can give a local measure of dependence between two random variables X_1 and X_2, see the discussion in Tjøstheim and Hufthammer (2013) for further details, and it is (as will be discussed in section 1.2) possible to use these local measures of dependence to construct different local approaches to spectral analysis.

Some local measures of the dependency structures between X_1 and X_2 contains expressions based on entities like $X_1 \leq p_1$ and $X_2 \leq p_2$, where (p_1, p_2) denotes a point of interest.[1] The knowledge of the distribution function F° of the bivariate pair (X_1, X_2) will in general be unknown, and a transformation from observations to pseudo-uniform observations (using part I of eq. (1.4)) is thus common to apply before the estimation of these local measures. The transformation of the point (p_1, p_2) is a pair of quantiles (τ_1, τ_2), and it is thus natural to refer to the local measures obtained in this manner as *quantile based*.

It can be instructive to give a simple example that combines a visualisation of the transformation process from eq. (1.4) with a brief comparison of the way the local Gaussian correlation ρ_v and the quantile based local measures are estimated. For this purpose consider the lag 1 pairs $\left(Y_{t+1}, Y_t\right)$ from a sample of size 400 from an AR(2)-model

[1] Variants based on expression like $p_1' < X_1 \leq p_1$ can also be encountered – but those are for simplicity not included in the present discussion.

$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + w_t$, with $\phi_1 = 1$, $\phi_2 = -0.75$ and $w_t \sim \mathrm{N}(0, 1)$.

The panels of fig. 1.2 shows the 399 lag 1 pairs $\left(Y_{t+1}, Y_t\right)$ based on the original observations, the pseudo-uniform observations and the pseudo-normalised observations, of which the latter two respectively can be used for the estimation of a quantile based measure of dependence and a local Gaussian measure of dependence.

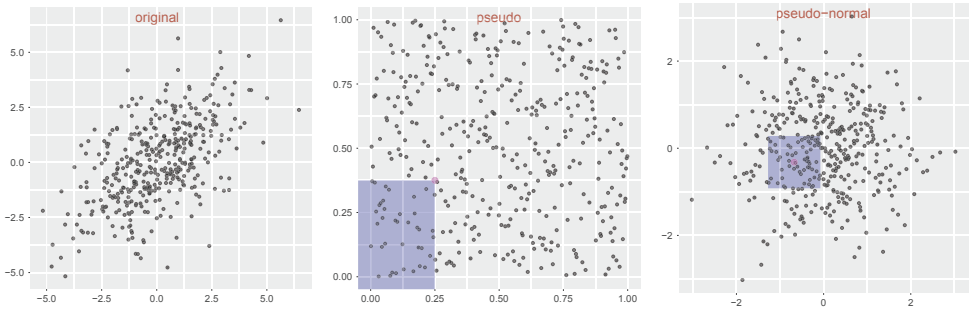

Figure 1.2: The 399 lag 1 pairs $\left(Y_{t+1}, Y_t\right)$ from the AR(2)-example. Left: Original observations. Center: Pseudo-uniform observations. Right: Pseudo-normalised observations.

The blue area in the center panel shows the pseudo-uniform observations that, given the quantiles $\boldsymbol{q} = (0.25, 0.375)$, will be essential for the estimation of a quantile based measure of dependence for that particular quantile. The point $\boldsymbol{v} = (-0.67, -0.32)$ in the right panel is the Φ^{-1}-transformed version of $\boldsymbol{q}$, and the blue square centered at $\boldsymbol{v}$ (based on the bandwidth $\boldsymbol{b} = (0.6, 0.6)$) shows the pseudo-normalised observations that are essential for the estimation of the local Gaussian correlation in this case.

The blue regions in fig. 1.2 highlights the pseudo-uniform observations and pseudo-normalised observations that will be essential for the estimation of the local measures of dependence for the given sample, and it is important to note that these regions behave differently when the sample size changes; in particular, the region in the center panel will remain the same regardless of the sample size, whereas the square in the right panel will shrink when the sample size increases (since a smaller bandwidth $\boldsymbol{b}$ then can be applied).

Note that in theory it is possible to compute the local Gaussian correlations based on the original observations in the left panel or the pseudo-uniform observations in the center panel, but that this in practice might not be advisable to do due to numerical problems in the estimation algorithm.

1.2 Theoretical framework: Time series

1.2.1 Stationary time series and spectral densities

The discussion will for notational simplicity focus on the univariate time series. The multivariate case will be mentioned briefly in the discussion of Paper 2.

For a univariate time series $\{Y_t\}_{t\in\mathbb{Z}}$, let $\gamma(r,s)$ denote the covariance between the terms Y_r and Y_s. The time series is weakly stationary if it satisfies the following three requirements

$$\text{(i) } \mathrm{E}\left[|Y_t|^2\right] < \infty \quad \text{for all } t \in \mathbb{Z}, \tag{1.5a}$$

$$\text{(ii) } \mathrm{E}[Y_t] = m \quad \text{for all } t \in \mathbb{Z}, \tag{1.5b}$$

$$\text{(iii) } \gamma(r,s) = \gamma(r+t, s+t) \quad \text{for all } r,s,t \in \mathbb{Z}. \tag{1.5c}$$

In this case the covariance only depends on the difference $h = r - s$, and it is then denoted $\gamma(h)$, *the autocovariance at lag* h. The corresponding entity $\rho(h) = \gamma(h)/\gamma(0)$ is *the autocorrelation at lag* h. If the collection $\{\gamma(h)\}_{h\in\mathbb{Z}}$ is absolutely summable, then the spectral density $f(\omega)$ is the Fourier transform of $\{\gamma(h)\}_{h\in\mathbb{Z}}$, i.e.

$$f(\omega) \coloneqq \sum_{h\in\mathbb{Z}} \gamma(h) \cdot e^{-2\pi i\omega h}. \tag{1.6}$$

It follows from the Spectral Representation Theorem, see e.g. Brockwell and Davis (1986, Th. 4.8.2), that it exists a right continuous orthogonal-increment processes $Z(\omega)$ that (with probability one) enables the original stationary time series to be expressed as

$$Y_t = \int_{-1/2}^{1/2} e^{2\pi i t\omega} \,\mathrm{d}\, Z(\omega), \tag{1.7}$$

which can be considered as a regression (with stochastic coefficients) of the time series $\{Y_t\}_{t\in\mathbb{Z}}$ on sines and cosines that oscillates at different frequencies.

It can from this be derived that the spectral density $f(\omega)$, when it exists, satisfies

$$f(\omega)\,\mathrm{d}\omega = \mathrm{E}\left[|\mathrm{d}\, Z(\omega)|^2\right], \tag{1.8}$$

i.e. $f(\omega)$ is related to the variance of the complex-valued entity $\mathrm{d}\, Z(\omega)$.

Furthermore, the inverse Fourier transform of eq. (1.6) gives the relation

$$\gamma(h) = \int_{-1/2}^{1/2} f(\omega) \cdot e^{2\pi i\omega h} \,\mathrm{d}\omega, \tag{1.9}$$

which for $h = 0$ expresses the variance $\mathrm{Var}(Y_t) = \gamma(0)$ as the integral of the spectral density $f(\omega)$. This enables a visual inspection of how much different frequencies contributes to the variance, and peaks and troughs in the graph of $f(\omega)$ can thus reveal information about periodic properties of the time series $\{Y_t\}_{t\in\mathbb{Z}}$.

Figure 1.3 exemplifies this for the case of the AR(2)-model from section 1.1.6. The panel at the bottom shows the theoretical spectral density $f(\omega)$ (only plotted for $\omega \geq 0$ since $f(\omega)$ is an even function), and the peak near 0.15 indicates that a sample from this AR(2)-model should have some periodic behaviour. However, since the peak is rather wide, it will not be a clear-cut periodicity that occurs, which is in agreement with the

observed trace shown in the top panel.

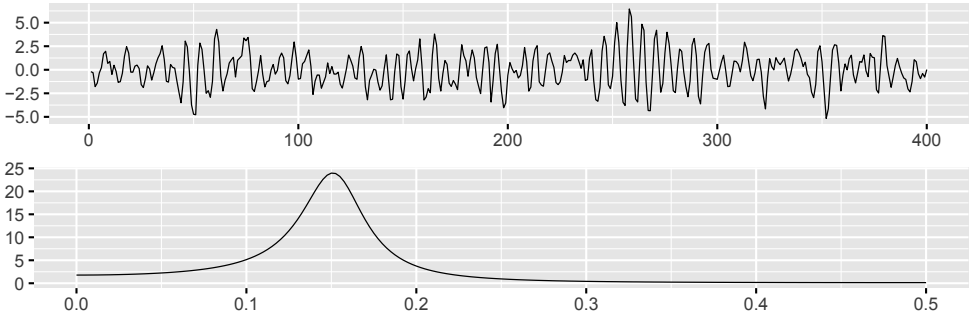

Figure 1.3: An AR(2)-example. Top panel: Trace. Bottom panel: $f(\omega)$.

Since Gaussian distributions are completely determined by their first and second moments, it follows from eq. (1.5) that there is a one-to-one correspondence between Gaussian time series and their corresponding spectral densities. The situation is not that nice for general time series, and $f(\omega)$ does in particular fail to be useful for GARCH-type models, where the observations $\{Y_t\}_{t\in\mathbb{Z}}$ are uncorrelated but not independent. The spectral density is constant for such GARCH-type models, and $f(\omega)$ thus does not provide any information about the underlying nonlinear structure.

Many approaches have been investigated in order to find generalisations of $f(\omega)$ that overcomes the above mentioned shortcomings. Some of them (like the new approach presented in this thesis) need the additional assumption that the time series $\{Y_t\}_{t\in\mathbb{Z}}$ should be *strictly stationary*, i.e. that the joint distributions of $\left[Y_{t_1}, \dots Y_{t_k}\right]'$ and $\left[Y_{t_1+h}, \dots Y_{t_k+h}\right]'$ are the same for all positive integers k and for all $t_1, \dots, t_k, h \in \mathbb{Z}$.

Before presenting the *local Gaussian auto- and cross-spectra* in Paper 1 and 2, it can be instructive first to give a brief account of some of the existing global and local generalisations of $f(\omega)$.

1.2.2 Higher order spectra (global) generalisations of $f(\omega)$

Noting that eq. (1.9), for $h = 0$, gave a decomposition of the variance (second order moment) over different frequencies, Tukey introduced similar constructions based on higher order moments and cumulants,[2] e.g. the bispectrum and trispectrum, which respectively can be considered as a decomposition of skewness and kurtosis over the frequencies – and he also introduced the polyspectra for multivariate time-series. See Brillinger (1991) for an overview of the early history of the higher-order spectra.

[2]This approach was mentioned in Lapierre and Fortet (1953), and according to Brillinger (2002, p. 1602) this could have motivated Tukeys 1953 article *The spectral representation and transformation properties of the higher moments of a stationary time series*, in which the definition and properties of the bispectrum (the third-order spectral density) was investigated. This article was unpublished until it appeared in *The Collected Works of John W. Tukey*, Brillinger (1984), and the bispectrum first appeared in print in Tukey (1959).

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 occur for the other generalisations too, and the following quote from Akaike (1966) should thus be kept in mind 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.

Despite these issues, higher-order spectra have been further investigated and applied in a plethora of different cases – the overview article Swami et al. (1997) lists 1759 entries up to 96/97, and a similar counting today would probably dwarf that number. This increase is related to the improvement of the available computational and graphical resources, which enables researchers to investigate nonlinear time series by means of these higher order spectra techniques.

For further details, the interested reader could e.g. consult Collis et al. (1998), which discuss the higher cumulant versus higher moment approaches, highlight technical details related to the estimation-techniques, and which, in addition, presents some approaches with regard to the graphical inspection of complex-valued results.

1.2.3 Alternative (global) generalisations of $f(\omega)$

The idea of the *higher order spectra* approach was to extend $f(\omega)$ by replacing $\gamma(h)$ (a second order moment measure) with measures that took into account higher moments or cumulants. However, as discussed in Van Hecke et al. (2017), it is possible to consider general spectral densities of the form

$$f_{\xi}(\omega) = \sum_{h \in \mathbb{Z}} \xi_h \cdot e^{-2\pi i \omega h}, \tag{1.10}$$

where the quantity ξ_h is some dependence measure computed from the random variables Y_t and Y_{t+h}. As for instance noted in (Van Hecke et al., 2017), generalisations of $f(\omega)$ of the form given in eq. (1.10) have been considered in Ahdesmäki et al. (2005) (Kendall's τ), Carcea and Serfling (2015) (Gini autocovariance) and Zhou (2012) (L-moments).

Note that the local Gaussian approach to spectral analysis is based on the idea described in eq. (1.10), but with the variation that ξ_h will depend on the selection of a point $\boldsymbol{v} = (v_1, v_2)$. This enables an investigation of how different strata of the time series interact, and for points on the diagonal, i.e. $v_1 = v_2$, it might 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.

1.2.4 Generalised function (global) generalisation of $f(\omega)$

Hong (1999) introduced a generalisation of $f(\omega)$ based on the characteristic function. The idea is to transform a stationary time series $\{Y_t\}_{t\in\mathbb{Z}}$ into a time series based on the corresponding characteristic functions, i.e. $\{\exp(iuY_t)\}_{t\in\mathbb{Z}}$, and then consider the bivariate function $\sigma_j(u,v) \coloneqq \mathrm{Cov}\big(\exp(iuY_t), \exp\big(ivY_{t-|j|}\big)\big)$. 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

$$h(\omega,u,v) \coloneqq \sum_{h\in\mathbb{Z}} \sigma_j(u,v)\cdot e^{-2\pi ij\omega}, \tag{1.11}$$

All the information about $\{Y_t\}_{t\in\mathbb{Z}}$ is contained in eq. (1.11), the ordinary spectrum can e.g. be restored as $f(\omega) = -\frac{1}{\sigma^2}\frac{\partial^2}{\partial u\partial v}h(\omega,u,v)\Big|_{(u,v)=(0,0)}$. 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.

1.2.5 Quantile-based (local) generalisations of $f(\omega)$

It is well known that stock returns behave in an asymmetric manner, i.e. that they, as noted in Hong et al. (2007), 'more often move with the market when the market goes down than when it goes up'. This asymmetry can not be detected by a (global) measure like the autocorrelation $\gamma(h)$, and several *local approaches* have been developed in order to help detect these asymmetries. After the development of a local replacement for the autocovariances $\gamma(h)$, it is then natural to Fourier transform them, like in eq. (1.10), in order to obtain a corresponding *local spectral density*.

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

$$\gamma_h(x_1,x_2) \coloneqq \mathrm{Cov}\big(\mathbb{1}\{Y_{t+h}\le x_1\}, \mathbb{1}\{Y_t\le x_2\}\big), \qquad (x_1,x_2)\in\mathbb{R}^2, \tag{1.12a}$$

$$\gamma_h^U(\tau_1,\tau_2) \coloneqq \mathrm{Cov}\big(\mathbb{1}\{U_{t+h}\le \tau_1\}, \mathbb{1}\{U_t\le \tau_2\}\big), \qquad (\tau_1,\tau_2)\in(0,1)^2, \tag{1.12b}$$

where $\mathbb{1}\{\cdot\}$ is the indicator function and where knowledge of the marginal distribution G is necessary in order to construct $U_t \coloneqq G(Y_t)$. 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.)

The copula spectral density kernel in Dette et al. (2015) is closely related to the concept of *quantile regression*, introduced in Koenker and Bassett Jr (1978), see also Koenker (2005). Several other local alternatives to the ordinary spectral density $f(\omega)$ 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. A *quantile periodogram* was investigated in Li (2012c) (see also Li (2008, 2010a,b,c, 2012a,b, 2014a)), and a quantile-based approach was also applied in Hagemann (2011).

Note that not all of these approaches are of the form given in eq. (1.10), e.g. the *quantile periodogram* from Li (2012c), and it might thus be natural to include some remarks from Li (2014b, p. 252-253)) about the construction of the *quantile periodogram* from a sample $\{y_t\}_{t=1}^n$. The starting point is the function $\rho_\alpha(x) \coloneqq x \cdot \{\alpha - \mathbb{1}\{x < 0\}\}$, which is $(\alpha - 1)x$ when $x < 0$ and αx when $x \geq 0$. This is used in the estimation of two different quantile-estimates, of which the first estimate is the sample α-quantile, denoted by $\hat{\lambda}_\alpha$, which minimises $\sum_{t=1}^n \rho_\alpha(y_t - \tilde{\lambda})$. The other estimate is based on a *quantile regression* of the α-quantile of y_t, which is the λ_t that satisfies $\mathrm{P}(y_t \leq \lambda_t) = \alpha$. The idea in this latter case is that λ_t can be estimated by $\tilde{\lambda} + \boldsymbol{x}_t'(\omega)\boldsymbol{\beta}$, where $\boldsymbol{x}_t(\omega)$ is the trigonometric regressor $[\cos(2\pi\omega t), \sin(2\pi\omega t)]'$, and where the optimal regression coefficients $\hat{\lambda}_\alpha(\omega)$ and $\hat{\boldsymbol{\beta}}_\alpha(\omega)$ are found by minimising $\sum_{t=1}^n \rho_\alpha(y_t - \tilde{\lambda} - \boldsymbol{x}_t'(\omega)\boldsymbol{\beta})$. The *quantile periodogram* is then defined by

$$Q_{n,\alpha}(\omega) \coloneqq \sum_{t=1}^n \rho_\alpha\big(y_t - \hat{\lambda}_\alpha\big) - \sum_{t=1}^n \rho_\alpha\big(y_t - \hat{\lambda}_\alpha(\omega) - \boldsymbol{x}_t'(\omega)\hat{\boldsymbol{\beta}}_\alpha(\omega)\big)\,, \tag{1.13}$$

which gives a measure of the net contribution of the trigonometric regressor $\boldsymbol{x}_t'(\omega)$ to the reduction of the quantile regression cost function.

1.3 The local Gaussian spectral densities

1.3.1 The local Gaussian auto-spectrum, definition and basic properties

The asymmetric nature of e.g. stock returns, see the discussion in section 1.2.5, can also be investigated using the local Gaussian approach – and as for all the other local measures of dependence, the natural question is then to examine whether a *local spectral analysis* can be based on this. It is in particular of interest to investigate if it might be possible to detect periodicities/cycles at a *local scale* (e.g. in the upper and lower extremes of the time series under investigation) which are invisible to the ordinary spectral density $f(\omega)$ that works on the *global scale*.

For a point $\boldsymbol{v} = (v_1, v_2)$ and a strictly stationary time series $\{Y_t\}_{t\in\mathbb{Z}}$, the construction of the *local Gaussian auto-spectrum* $f_{\boldsymbol{v}}(\omega)$, requires that it is possible to compute the

collection of *local Gaussian autocorrelations* $\rho_{\boldsymbol{v}}(h)$ for all $h \neq 0$, i.e. that all the probability density functions g_h for the bivariate random variables $\left(Y_{t+h}, Y_t\right)$, satisfies the regularity conditions given in Tjøstheim and Hufthammer (2013). The case $h = 0$ can not be covered by this, but for this case it is natural to let $\rho_{\boldsymbol{v}}(0)$ be identical to 1 for all $\boldsymbol{v} \in \mathbb{R}^2$.

If the collection $\{\rho_{\boldsymbol{v}}(h)\}_{h\in\mathbb{Z}}$ is absolutely summable, i.e. $\sum_{h\in\mathbb{Z}} |\rho_{\boldsymbol{v}}(h)| < \infty$, then the *local Gaussian auto-spectrum* can be defined as the Fourier transform of $\{\rho_{\boldsymbol{v}}(h)\}_{h\in\mathbb{Z}}$, i.e.

$$f_{\boldsymbol{v}}(\omega) \coloneqq \sum_{h\in\mathbb{Z}} \rho_{\boldsymbol{v}}(h) \cdot e^{-2\pi i\omega h}. \tag{1.14}$$

The estimation problems mentioned in section 1.1.5 motivate an adjusted definition, see Paper 1, where an initial step *normalises* the time series under investigation, i.e. $\{Y_t\}_{t\in\mathbb{Z}}$ will be replaced with $\{Z_t \coloneqq \Phi^{-1}\left(G\left(Y_t\right)\right)\}_{t\in\mathbb{Z}}$, where G is the cumulative density function[3] of Y_t and Φ^{-1} as usual is the inverse of the cumulative density function of the standard normal distribution. It will henceforward for simplicity be assumed that this transformation has been performed when $f_{\boldsymbol{v}}(\omega)$ is discussed.

For Gaussian time series, the *local Gaussian autocorrelations* $\rho_{\boldsymbol{v}}(h)$ coincide with the ordinary (global) autocorrelations $\rho(h)$, which implies that $f_{\boldsymbol{v}}(\omega)$ in that case coincides with $f(\omega)/\gamma(0)$, the variance-rescaled version of the ordinary spectral density from eq. (1.6). The assumption that the investigation is performed on the normalised version of the time series implies that $\gamma(0) = 1$, so $f_{\boldsymbol{v}}(\omega) = f(\omega)$ for all points $\boldsymbol{v} \in \mathbb{R}^2$ when the time series under investigation is Gaussian.

It follows from this that a comparison of $f_{\boldsymbol{v}}(\omega)$ and $f(\omega)$ can be used to identify non-Gaussian traits in the time series under investigation. If $f(\omega)$ is flat, i.e. $\rho(h) = 0$ for all $h \neq 0$, then a non-flat $f_{\boldsymbol{v}}(\omega)$ can be taken as an indicator of non-linear dependency structures in the time series under investigation, and the peaks and troughs of $f_{\boldsymbol{v}}(\omega)$ might then reveal additional information about these non-linear structures.

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 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 the coincidence of $f_{\boldsymbol{v}}(\omega)$ and $f(\omega)$ for Gaussian time series is thus quite desirable from this point of view.

A comparison of $f_{\boldsymbol{v}}(\omega)$ and $f(\omega)$ for a general point $\boldsymbol{v} = (v_1, v_2)$ can be a bit complicated, since the *local Gaussian autocorrelations* $\rho_{\boldsymbol{v}}(h)$ in general are different from $\rho_{\boldsymbol{v}}(-h)$, which implies that $f_{\boldsymbol{v}}(\omega)$ then becomes a complex valued function. That issue is not present if the point $\boldsymbol{v}$ lies on the diagonal, i.e. $v_1 = v_2$, so the examples in the next section will only consider such diagonal points where $f_{\boldsymbol{v}}(\omega)$ is a real valued function.

[3] In practice G will be unknown, and an estimate $\widehat{G}$ (based on the sample at hand) will be used instead. This adjustment adds some smaller order effects that must be taken into account during the theoretical analysis.

1.3.2 Estimation of $f_{\boldsymbol{v}}(\omega)$ and some examples

Theoretical and numerical estimates of the ordinary spectral density $f(\omega)$ is typically investigated by means of the fast Fourier transform (FFT) and techniques related to the periodogram. The effectiveness of the FFT is due to the product structure of the estimated autocovariances, i.e. $\widehat{\gamma}(h) = \sum_{t=1}^{n-|h|} y_{t+|h|} \cdot y_t$, and since no similar product structure is present for the estimates used in the local Gaussian approach (which is based on the distribution), it is not feasible to use the techniques from the FFT-based theory for estimates of $f(\omega)$ when estimates of $f_{\boldsymbol{v}}(\omega)$ are required.

The pre-FFT approach for estimation of $f(\omega)$, where a Fourier transform is taken of the estimated autocorrelations after they have been smoothed and truncated by means of some lag-window function, can be adapted to deal with the estimates of the local Gaussian spectral densities $f_{\boldsymbol{v}}(\omega)$. The estimates to be used later on for the comparison of $f(\omega)$ and $f_{\boldsymbol{v}}(\omega)$ will thus be

$$\widehat{f}^{m}(\omega) = \sum_{|h|\leq m} \lambda_{m}(h) \cdot \widehat{\rho}(h) \cdot e^{-2\pi i\omega h}, \tag{1.15a}$$

$$\widehat{f}^{m}_{\boldsymbol{v}}(\omega) = \sum_{|h|\leq m} \lambda_{m}(h) \cdot \widehat{\rho}_{\boldsymbol{v}}(h) \cdot e^{-2\pi i\omega h}, \tag{1.15b}$$

where m denotes the truncation level and $\lambda_{m}(h)$ is the lag-window function used for the smoothing. It has not been explicitly included in the notation of eq. (1.15b), but the estimate $\widehat{f}^{m}_{\boldsymbol{v}}(\omega)$ does also depend on the bandwidth $\boldsymbol{b}$ that is needed for the estimation of the local Gaussian autocorrelations $\rho_{\boldsymbol{v}}(h)$.

The estimation of $f_{\boldsymbol{v}}(\omega)$, based on a (pseudo-normalised) sample $\{y_t\}_{t=1}^{n}$ of size n, must also take into account that the selection of the point $\boldsymbol{v} = (v_1, v_2)$ will influence the bias-variance properties of the estimate. In particular, it can be problematic if the coordinates v_1 and v_2 are too far out in the tails of the sample. And moreover, as mentioned in the previous section, it can be preferable to consider points on the diagonal, i.e. $v_1 = v_2$, since $f_{\boldsymbol{v}}(\omega)$ (and the estimate $\widehat{f}^{m}_{\boldsymbol{v}}(\omega)$) then are real valued functions of ω. The coordinates v_1 and v_2 will, for the examples investigated in Paper 1 and 2, be chosen among the 10%, 50% and 90% percentiles of the standard normal distribution, whose values are -1.28, 0 and 1.28. Information about $\boldsymbol{v}$ is contained in the upper right corner of the relevant plots, where it is marked as `10%::10%`, and so on.

Figures 1.4 and 1.5 are taken from Paper 1, and show two simulation based investigations of $f_{\boldsymbol{v}}(\omega)$. Figure 1.4 shows the result when the estimation procedure is used on 100 independent samples from a standard normal distribution $\mathrm{N}(0,1)$. 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 correspond to the coordinates v_1 and v_2. The three other panels contain 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 $f^{m}_{\boldsymbol{v}}(\omega)$ for the three diagonal points under investigation (blue part). Information about the truncation level and the points are printed at the top of each plot.

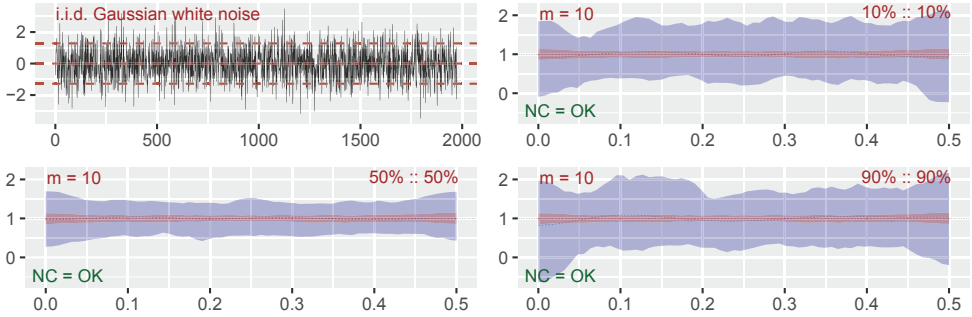

Figure 1.4: i.i.d. Gaussian white noise

It can be seen from fig. 1.4 that the medians of the estimates (the red and blue dashed lines) are good estimates of $f^m(\omega)$ and $f_v^m(\omega)$ (the m-truncations of the true values), which in this case in fact coincides with $f(\omega)$ and $f_v(\omega)$, 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 make the estimates more prone to small-sample variation. Note also that the confidence bands are wider when the coordinates v_1 and v_2 are in the tails, which is as expected due to the reduced number of available observations.

The estimation procedure gave good estimates of the true values $f(\omega)$ and $f_v(\omega)$ in the simple example of fig. 1.4, but it is important to keep in mind that these plots actually shows estimates of $f^m(\omega)$ and $f_v^m(\omega)$. It might be necessary to apply a (much) higher truncation level before $f^m(\omega)$ and $f_v^m(\omega)$ give decent approximations of the true values $f(\omega)$ and $f_v(\omega)$. It thus seems preferable to compute $\widehat{f_v^m}(\omega)$ for a range of possible truncation levels m, and then check if the shapes of the estimates for different truncations share the same properties with regard to the position of any peaks and troughs. See the discussion of the computer code in section 1.4 for further details.

The Gaussian case investigated in fig. 1.4 is the only one where it is known what the true value of $f_v(\omega)$ should be. However, it is possible to construct an artificial example like the one investigated in fig. 1.5, where an argument can be given that justifies that the three *local peaks* of $f_v^m(\omega)$ are as expected based on the generating model and the bandwidth $\boldsymbol{b} = (0.6, 0.6)$ used in this particular case. Note that the parameters used to generate the artificial example in fig. 1.5 was selected such that the global spectrum $f(\omega)$ should look like white noise, and this is in agreement with the observed estimates of the global spectrum.

Note that fig. 1.5 shows that $f_v(\omega)$, in contrast to $f(\omega)$, can take negative values for some frequencies. This is due to the fact that the collection of local Gaussian autocorrelations $\{\rho_v(h)\}_{h\in\mathbb{Z}}$ might not be non-negative definite. Furthermore, a correspondence like eq. (1.8) does not exist for $f_v(\omega)$, and caution is thus advised if peaks and troughs of $f_v(\omega)$ are attempted interpreted exactly as they would have been if they had occurred

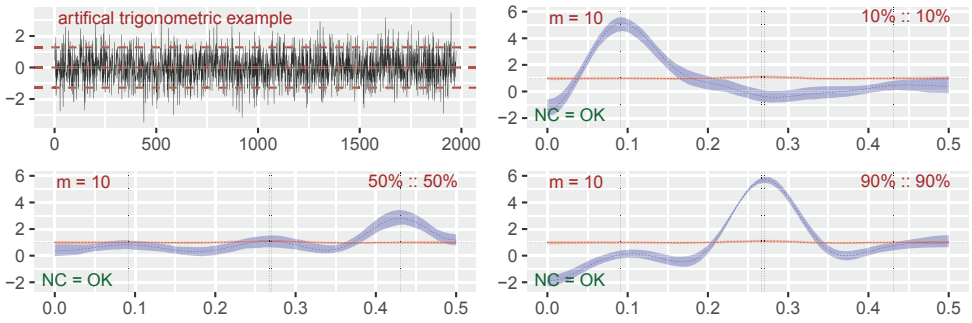

Figure 1.5: Artifical example, *hidden trigonometric components*.

for the ordinary spectral density $f(\omega)$.

These somewhat problematic issues for $f_{\boldsymbol{v}}(\omega)$ might for practical purposes be a minor problem, i.e. the central part of interest is that $f_{\boldsymbol{v}}(\omega)$ can detect if a time series $\{Y_t\}_{t\in\mathbb{Z}}$ contains non-Gaussian traits. And, as was seen in fig. 1.5, the shape of $f_{\boldsymbol{v}}(\omega)$ can reveal local properties that goes unnoticed by the ordinary spectral density. These features implies that estimates of the local Gaussian spectral densities might be of interest to consider as an additional tool with regard to local frequency representation and model interpretation for time series, a topic that is discussed in Paper 1 and 2.

1.3.3 The local Gaussian cross-spectrum

The cross-spectrum $f_{k\ell}(\omega)$ is a bivariate concept defined for a jointly stationary time series $\left\{\left(Y_{k,t}, Y_{\ell,t}\right)\right\}_{t\in\mathbb{Z}}$, which in Paper 2 will be assumed to consists of two of the margins of a stationary multivariate time series $\{\boldsymbol{Y}_t\}_{t\in\mathbb{Z}}$, where $\boldsymbol{Y}_t = \left(Y_{1,t}, \ldots, Y_{d,t}\right)$. With $\rho_{k\ell}(h) \coloneqq \operatorname{Cov}\left(Y_{k,t+h}, Y_{\ell,t}\right)$, the standardised cross-spectrum can be defined as

$$f_{k\ell}(\omega) \coloneqq \sum_{h\in\mathbb{Z}} \rho_{k\ell}(h) \cdot e^{-2\pi i\omega h}, \tag{1.16}$$

and the *local Gaussian cross-spectrum* at a point $\boldsymbol{v} = (v_1, v_2)$ is then simply given as

$$f_{k\ell:\boldsymbol{v}}(\omega) \coloneqq \sum_{h\in\mathbb{Z}} \rho_{k\ell:\boldsymbol{v}}(h) \cdot e^{-2\pi i\omega h}, \tag{1.17}$$

i.e. ordinary cross-covariances are replaced with the *local Gaussian cross-covariances* that corresponds to the density functions $g_{k\ell}$ of the bivariate random variables $\left(Y_{k,t+h}, Y_{\ell,t}\right)$.

The cross-spectrum $f_{k\ell}(\omega)$ can, based on the same reasoning that was used in eqs. (1.7) and (1.8), be expressed as

$$f_{k\ell}(\omega)\,\mathrm{d}\omega = \mathrm{E}\left[\left|\mathrm{d}\,Z_k(\omega)\overline{\mathrm{d}\,Z_\ell(\omega)}\right|\right], \tag{1.18}$$

where $Z_k(\omega)$ and $Z_\ell(\omega)$ are the right continuous orthogonal-increment processes that

correspond to $\{Y_{k,t}\}_{t\in\mathbb{Z}}$ and $\{Y_{\ell,t}\}_{t\in\mathbb{Z}}$, i.e. $f_{k\ell}(\omega)$ is related to the covariance of the two complex-valued entities $\mathrm{d}\,Z_k(\omega)$ and $\mathrm{d}\,Z_\ell(\omega)$.

The local Gaussian cross-spectrum $f_{k\ell:\boldsymbol{v}}(\omega)$ will by construction coincide with $f_{k\ell}(\omega)$ if the multivariate time series under investigation is Gaussian, and the discussions related to the univariate time series thus apply in the multivariate case too.

Since $f_{k\ell}(\omega)$ in general is a complex-valued function, the actual comparison with $f_{k\ell:\boldsymbol{v}}(\omega)$ will be based on the original and local Gaussian versions of the cospectrum, quadrature spectrum, phase spectrum and amplitude spectrum.

Note that the *coherence*, defined as the correlation of $\mathrm{d}\,Z_k(\omega)$ and $\mathrm{d}\,Z_\ell(\omega)$, and the related concept of *squared coherence* (a number between 0 and 1), does *not* behave well in the local Gaussian case. The squared coherence can be expressed as $f_{k\ell}(\omega)f_{\ell k}(\omega)/f_{kk}(\omega)f_{\ell\ell}(\omega)$, where a repeated index indicates an auto-spectrum, and it is possible to create a *naive* local Gaussian analogue of these concepts, by simply replacing the ordinary spectra with their local Gaussian spectra. This naive generalisation of the squared coherence was however dropped, since it in general returned complex valued results without any restrictions of how large the amplitudes could become.

1.3.4 Local Gaussian spectral analysis versus other alternatives to $f(\omega)$

It has not been the aim of this thesis to compare the local Gaussian approach to spectral analysis with any of those approaches mentioned in sections 1.2.2 to 1.2.5, but a few comments can nevertheless be included to at least point upon some details of interest.

The main motivation for these approaches is to investigate those properties of a time series $\{Y_t\}_{t\in\mathbb{Z}}$ that are not captured by the second order moments, which in particular is of interest for non-Gaussian time series where nonlinear structures might be present. The local Gaussian approach requires the assumption of strict stationarity, since it is based on a local measure defined relative to the distribution functions, and this requirement is shared by many of the approaches mentioned in sections 1.2.4 and 1.2.5.

The copula parts of the different distributions are central for the local Gaussian approach, due to the practical requirement that the estimates are computed based on pseudo-normalised observations. The additional normalising part of the transformation makes this slightly different from the copula-approach described in Dette et al. (2015), but it does still separate the marginal aspect of the time series from the serial aspect. Moreover, as explained in Dette et al. (2015), the step that replaces the original observations with pseudo-uniform observations implies that these methods are invariant with respect to monotone transformations of the marginal density, e.g. the same result will be obtained from $\{Y_t\}_{t\in\mathbb{Z}}$ and $\{\exp(Y_t)\}_{t\in\mathbb{Z}}$.

1.3.5 Local Gaussian spectral analysis and model selection

Issues related to model selection were discussed in section 1.1.2, and the selection of copula models will in particular be looked upon in Paper 3. Model selection can also

be of interest to consider in conjunction with the local Gaussian spectral densities, as explained in the discussion of the financial time series examples in Paper 1 and 2.

The basic idea in this latter case is that, given some sample $\{y_t\}_{t=1}^n$ from a time series, and given a collection of potential models $\mathcal{G} := \{G_i\}_{i=1}^s$, then existing techniques should first be applied for the selection of the optimal parameters for each individual G_i and an overall best model $G^\star$ from $\mathcal{G}$ could then be selected for $\{y_t\}_{t=1}^n$.

An informal additional step can then be added based on the local Gaussian approach to spectral analysis. In particular, it is possible to compare the results based on the original sample $\{y_t\}_{t=1}^n$ with results obtained from samples generated from the fitted models. This might provide visual cues with regard to the suitability of the different models, e.g. if the models in a decent manner managed to capture asymmetric properties of the original observations.

1.4 Computer code

An `R`-package `localgaussSpec` has been created that builds upon the `localgauss`-package (Berentsen et al., 2014a) and computes the local Gaussian spectral densities discussed in Paper 1 and Paper 2. This package can be installed into R (R Core Team, 2017) by the help of the `devtools`-package (Wickham and Chang, 2016):

```
library(devtools)
install_github("LAJordanger/localgaussSpec")
```

The scripts needed for the reproduction of the plots in Paper 1 and Paper 2 are contained in the `localgaussSpec`-package, and the interested reader can extract these scripts from the compiled package by the help of the function `LG_extract_scripts`. Note that these scripts also document how the different functions of the package work together, and they are thus also intended to be used as templates for those that would like to check out the functionality of the `localgaussSpec`-package.

Note that a `shiny`-application (Chang et al., 2017) is used to enable an interactive investigation of the resulting plots, so it is fairly easy to investigate how the result is affected by different configurations of the input-parameters. Read the scripts and the package-documentation for further details.

1.5 Summary of papers

1.5.1 Summary of Paper 1: 'Nonlinear spectral analysis via the local Gaussian correlation.'

This paper introduces the *local Gaussian spectral density* $f_v(\omega)$ as a *local Gaussian* adjustment of the ordinary spectral density $f(\omega)$, by simply replacing the ordinary auto-correlations $\rho(h)$ in the definition of $f(\omega)$ with their *local Gaussian* counterparts $\rho_v(h)$.

An estimator of $f_{\boldsymbol{v}}(\omega)$ is constructed, and an asymptotic normality result for the estimator is obtained by extending arguments from Tjøstheim and Hufthammer (2013).

Some simulated examples are investigated in order to check that the estimated local Gaussian spectral density behaves as expected when it is known what the results should look like. It is verified that the estimates of $f_{\boldsymbol{v}}(\omega)$ and $f(\omega)$ are close to each other for Gaussian time series, and it is also seen that *local periodic phenomena* can be detected by $f_{\boldsymbol{v}}(\omega)$ for an artificially constructed time series whose ordinary (global) spectrum $f(\omega)$ looks like white noise.

A financial data set is then investigated by means of the local Gaussian spectral density, and it is observed that local nonlinear features are present. Simulations from a GARCH-type model fitted to the data set is also investigated, and it is seen that estimates of $f_{\boldsymbol{v}}(\omega)$ can be used to get a *local Gaussian spectral comparison* of the fitted model against the original observations. This implies that estimates of $f_{\boldsymbol{v}}(\omega)$ could play a role as an *exploratory tool* when it comes to model selection.

Note that the examples in this paper can be investigated in an interactive manner in the accompanying `R`-package `localgaussSpec`.

1.5.2 Summary of Paper 2: 'Nonlinear cross-spectrum analysis via the local Gaussian correlation.'

This paper extends the setup from Paper 1 to the multivariate case, i.e. the concept of *local Gaussian cross-spectra* $f_{k\ell;\boldsymbol{v}}(\omega)$ are defined from the ordinary (global) cross-spectrum $f_{k\ell}(\omega)$ by replacing the cross-correlations $\rho_{k\ell}(h)$ with their local Gaussian counterparts $\rho_{k\ell;\boldsymbol{v}}(h)$. It is clear that some concepts related to the complex-valued ordinary (global) cross-spectrum $f_{k\ell}(\omega)$ can be extended trivially to the *local Gaussian case* $f_{k\ell;\boldsymbol{v}}(\omega)$, i.e. local Gaussian versions exist for the *cospectrum*, *quadrature spectrum*, *amplitude spectrum* and *phase spectrum*.

The asymptotic theory for estimates of $f_{k\ell;\boldsymbol{v}}(\omega)$ is in essence the same as the asymptotic theory for estimates of $f_{\boldsymbol{v}}(\omega)$ (from Paper 1), which simplifies the investigation tremendously. The only new part is that it in the multivariate context also is necessary to find asymptotic results for the estimates of the *local Gaussian amplitude spectrum* and the *local Gaussian phase spectrum*, but this follows quite trivially by applying standard techniques as given in Brockwell and Davis (1986).

The examples in Paper 2 closely resembles those from Paper 1, i.e. first a few bivariate time series are considered that verifies that the expected results are returned for some simple cases, in particular a *local trigonometric example* shows that local periodic phenomena can be detected that goes unnoticed by the ordinary (global) spectral densities. A multivariate financial time series with a corresponding GARCH-type model is also investigated in order to show that estimates of the *local Gaussian cross-spectrum* might be useful with regard to model interpretation and selection.

Note that the examples in this paper can be investigated in an interactive manner in the accompanying `R`-package `localgaussSpec`.

1.5.3 Summary of Paper 3: 'Model selection of copulas: AIC versus a cross validation copula information criterion.'

Lars Arne Jordanger and Dag Tjøstheim, 'Model selection of copulas: AIC versus a cross validation copula information criterion.', *Statistics and Probability Letters*, **92**, 249-255 (2014).

Paper 3 is motivated by Grønneberg and Hjort (2008, 2014), i.e. it contains the results from a simulation study that compares the small-sample performance of the *cross-validation Copula Information Criterion* (xv-CIC) from (Grønneberg and Hjort, 2014) with the performance of the ^{p}AIC, where ^{p}AIC denotes that the *Akaike Information Criterion* (from Akaike (1974)) is used on a (bivariate) sample where the (rescaled) empirical marginal cumulative density functions have been used to replace observations with pseudo-uniform observations.

Before (Grønneberg and Hjort, 2008) it was commonly assumed that the transformation from observations to pseudo-uniform observations should have no impact on the asymptotic theory, but that turned out to be an incorrect assumption. The asymptotic investigation in (Grønneberg and Hjort, 2008) resulted in two alternative copula information criteria, but those were of less practical interest since they were not generally applicable. The xv-CIC introduced in (Grønneberg and Hjort, 2014) did not suffer from that problem, and thus presented a (from an asymptotic point of view) more appropriate model selection tool for copula models than ^{p}AIC.

However: In a practical setting only a finite sample will be present, and the result from ^{p}AIC might then (due to the effect of small-sample variation) be just as good/bad as the results from the computationally much costlier xv-CIC.

The small-sample simulation study in Paper 3 shows that xv-CIC and ^{p}AIC in most cases will select the same model, and neither of them seems to be superior to the other one for the cases investigated.

Bibliography

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. 1.2.3

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 1.2.2

Akaike, H., Dec 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19 (6), 716–723.
URL http://ieeexplore.ieee.org/abstract/document/1100705/ 1.1.2, 1.5.3

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 1.1.4

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 1.1.4, 1.1.5, 1.4

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 1.1.4

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 1.1.4

Brillinger, D. R., 1965. An Introduction to Polyspectra. The Annals of Mathematical Statistics 36 (5), 1351–1374.
URL http://www.jstor.org/stable/2238424 1.3.1

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. 2

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 1.2.2

Brillinger, D. R., 2002. John W. Tukey's Work on Time Series and Spectrum Analysis. The Annals of Statistics 30 (6), 1595–1618.
URL http://www.jstor.org/stable/1558731 2

Brockwell, P. J., Davis, R. A., 1986. Time Series: Theory and Methods. Springer-Verlag New York, Inc., New York, NY, USA. 1.2.1, 1.5.2

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 1.2.3

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 1.4

Collis, W., White, P., Hammond, J., 1998. Higher-order spectra: the bispectrum and trispectrum. Mechanical Systems and Signal Processing 12 (3), 375–394.
URL http://www.sciencedirect.com/science/article/pii/S088832709790145X 1.2.2

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 1.2.5, 1.2.5, 1.3.4

Grønneberg, S., Hjort, N. L., 2008. The copula information criterion. Technical Report 7, Department of Matehmatics, University of Oslo. 1.1.3, 1.5.3

Grønneberg, S., Hjort, N. L., 2014. The Copula Information Criteria. Scandinavian Journal of Statistics 41 (2), 436–459.
URL http://dx.doi.org/10.1111/sjos.12042 1.1.3, 1.5.3

Hagemann, A., November 2011. Robust Spectral Analysis.
URL https://ssrn.com/abstract=1956581 1.2.5

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 1.2.5

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 1.1.4, 1.1.4

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 1.2.4, 1.2.4

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 1.2.4

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 1.2.5

Jordanger, L. A., Tjøstheim, D., 2017. Nonlinear spectral analsys via the local Gaussian correlation.
URL http://arxiv.org/abs/1708.02166v1 (document)

Koenker, R., 2005. Quantile Regression. Vol. 38 of Econometric Society Monographs. Cambridge University Press. 1.2.5

Koenker, R., Bassett Jr, G., 1978. Regression Quantiles. Econometrica 46 (1), 33–50.
URL http://www.jstor.org/stable/1913643 1.2.5

Kullback, S., Leibler, R. A., 03 1951. On Information and Sufficiency. Ann. Math. Statist. 22 (1), 79–86.
URL http://dx.doi.org/10.1214/aoms/1177729694 1.1.2

Lacal, V., Tjøstheim, D., 2016. Estimating and testing nonlinear local dependence between two time series, unpublished manuscript. 1.1.4

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 1.1.4

Lapierre, A. B., Fortet, R., 1953. Théorie des fonctions aléatoires: applications à divers phénomènes de fluctuation. Collection d'ouvrages de mathématiques à l'usage de physiciens. Masson et Cie.
URL https://books.google.no/books?id=9Sws-RTcuKkC 2

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 1.2.4

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 1.2.5

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/ 1.2.5

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/ 1.2.5

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 1.2.5

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/ 1.2.5

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 1.2.5

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 1.2.5

Li, T.-H., 2014a. Quantile Periodogram and Time-Dependent Variance. Journal of Time Series Analysis 35 (4), 322–340.
URL http://dx.doi.org/10.1111/jtsa.12065 1.2.5

Li, T.-H., 2014b. Time series with mixed spectra. CRC Press. 1.2.5

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 1.2.5

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 1.1.4

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 1.1.4

R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL https://www.R-project.org/ 1.4

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. 1.1.3

Støve, B., Tjøstheim, D., April 2014. Measuring asymmetries in financial returns: an empirical investigation using local Gaussian correlation. In: Haldrup, N., Meitz, M., Saikkonen, P. (Eds.), Essays in Nonlinear Time Series Econometrics. No. 9780199679959 in OUP Catalogue. Oxford University Press, pp. 307–329. 1.1.4

Støve, B., Tjøstheim, D., Hufthammer, K. O., 2014. Using local Gaussian correlation in a nonlinear re-examination of financial contagion. Journal of Empirical Finance 25, 62–82.
URL http://www.sciencedirect.com/science/article/pii/S0927539813000947 1.1.4

Swami, A., Giannakis, G. B., Zhou, G., 1997. Bibliography on higher-order statistics. Signal Processing 60 (1), 65 – 126.
URL http://www.sciencedirect.com/science/article/pii/S0165168497000650 1.2.2

Takeuchi, K., 1976. Distribution of informational statistics and a criterion of model fitting. Suri-Kagaku (Mathematical Sciences) 153, 12–18, in Japanese. 1.1.2

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 (document), 1.1.4, 1.1.6, 1.3.1, 1.5.1

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. 2

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 1.2.3, 1.2.3

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 1.2.4

Wickham, H., Chang, W., 2016. devtools: Tools to Make Developing R Packages Easier. R package version 1.10.0.
URL https://CRAN.R-project.org/package=devtools 1.4

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 1.2.3

Chapter 2

Papers

Paper I

2.1 Nonlinear spectral analysis via the local Gaussian correlation

Lars Arne Jordanger and Dag Tjøstheim

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 $\left(Y_{t+h}, Y_t\right)$ 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.

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

The inverse Fourier transform gives the relation $\gamma(h) = \int_{-1/2}^{1/2} f(\omega) \cdot e^{2\pi i \omega h} \, \mathrm{d}\omega$, 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* $f_{\boldsymbol{v}}(\omega)$ 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,

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

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 $\rho_{\boldsymbol{v}}(h)$ are used to define the *local Gaussian spectral density* $f_{\boldsymbol{v}}(\omega)$. 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 $\left(Y_{t+h}, Y_t\right)$ 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

[1] This 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 $Y_t = \int_{-1/2}^{1/2} e^{2\pi i t \omega} \, \mathrm{d}Z(\omega)$, 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 $f_{\boldsymbol{v}}(\omega)$.

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

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

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 $\sigma_j(u,v) := \mathrm{Cov}\big(\exp(iuY_t), \exp\left(ivY_{t-|j|}\right)\big)$. 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

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

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

$$\gamma_h(x_1,x_2) := \mathrm{Cov}\big(\mathbb{1}\{Y_{t+h}\le x_1\}, \mathbb{1}\{Y_t\le x_2\}\big), \qquad (x_1,x_2)\in\mathbb{R}^2, \tag{1.5a}$$

$$\gamma_h^U(\tau_1,\tau_2) := \mathrm{Cov}\big(\mathbb{1}\{U_{t+h}\le \tau_1\}, \mathbb{1}\{U_t\le \tau_2\}\big), \qquad (\tau_1,\tau_2)\in(0,1)^2, \tag{1.5b}$$

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 $\left(Y_{t+h}, Y_t\right)$ that depends on a point $\boldsymbol{v} = (v_1, v_2)$, 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* $\rho_{\boldsymbol{v}}(h)$ 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* $f_{\boldsymbol{v}}(\omega)$ 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* $f_{\boldsymbol{v}}(\omega)$ 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 $f_{\boldsymbol{v}}(\omega)$ can be used to detect and investigate nonlinear structures in non-Gaussian white noise, and in particular that $f_{\boldsymbol{v}}(\omega)$ 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 $f_{\boldsymbol{v}}(\omega)$. 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 strictly[3] 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 $\left(Y_{t+h}, Y_t\right)$.

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

For the purpose of this investigation, the vector containing the five parameters μ_1, μ_2, σ_1, σ_2 and ρ will be denoted by $\boldsymbol{\theta}$,[4] and the bivariate Gaussian density function will be denoted $\psi(\boldsymbol{w}; \boldsymbol{\theta})$, i.e.

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

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

$$g(\boldsymbol{v}) = \psi(\boldsymbol{w}; \boldsymbol{\theta})|_{\boldsymbol{w}=\boldsymbol{v}}, \qquad \frac{\partial}{\partial w_1} g(\boldsymbol{v}) = \frac{\partial}{\partial w_1} \psi(\boldsymbol{w}; \boldsymbol{\theta})\bigg|_{\boldsymbol{w}=\boldsymbol{v}}, \qquad \frac{\partial}{\partial w_2} g(\boldsymbol{v}) = \frac{\partial}{\partial w_2} \psi(\boldsymbol{w}; \boldsymbol{\theta})\bigg|_{\boldsymbol{w}=\boldsymbol{v}}, \tag{2.2}$$

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

It is easy to verify that a solution $\boldsymbol{\theta}$ can be found for any point $\boldsymbol{v}$ where $g(\boldsymbol{w})$ is smooth – but these solutions are not unique: $\psi(\boldsymbol{w}; \boldsymbol{\theta})$ and $\psi(\boldsymbol{w}; \boldsymbol{\theta}')$ can have coinciding first order linearisation around the point $\boldsymbol{v}$, 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 $\psi(\boldsymbol{w}; \boldsymbol{\theta})$ at the point $\boldsymbol{v}$, the approximation method needs to include a neighbourhood around $\boldsymbol{v}$. Applying the approach used when estimating densities in Hjort and Jones (1996), one can consider a $\boldsymbol{b} \to \boldsymbol{0}^+$ limit of parameters $\boldsymbol{\theta}_{\boldsymbol{b}}$ that minimise the penalty function

$$q = \int K_{\boldsymbol{b}}(\boldsymbol{w} - \boldsymbol{v}) \left[\psi(\boldsymbol{w}; \boldsymbol{\theta}) - g(\boldsymbol{w}) \log\left(\psi(\boldsymbol{w}; \boldsymbol{\theta})\right) \right] \mathrm{d}\boldsymbol{w}, \tag{2.3}$$

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

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

configuration $\boldsymbol{\theta}_{\boldsymbol{b}}$ for eq. (2.3) should solve the vector equation

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

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

$$\boldsymbol{\theta}_{0}=\lim_{\boldsymbol{b}\to\mathbf{0}^{+}}\boldsymbol{\theta}_{\boldsymbol{b}} \tag{2.5}$$

will be well-defined, and the ρ-part of the $\boldsymbol{\theta}_{0}$-vector can be used to define a LGC at the point $\boldsymbol{v}$.

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

$$\boldsymbol{W}\sim \mathrm{N}\left(\begin{bmatrix}\mu_{1}\\ \mu_{1}\end{bmatrix},\begin{bmatrix}\sigma_{1}^{2} & \sigma_{1}\sigma_{2}\rho\\ \sigma_{1}\sigma_{2}\rho & \sigma_{2}^{2}\end{bmatrix}\right), \tag{2.6}$$

then, for any point $\boldsymbol{v}$ and any bandwidth $\boldsymbol{b}$, the parameters $\boldsymbol{\theta}_{\boldsymbol{b}}$ 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 $\boldsymbol{v}$ 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 $G(\boldsymbol{w})$ can be expressed as $C(G_{1}(w_{1}),G_{2}(w_{2}))$, with $G_{i}(w_{i})$ the marginal cdf corresponding to W_{i}. This copula C contains all the interdependence information between the two marginal random variables W_{1} and W_{2}, it will be unique when the two margins are continuous, and it will then be invariant under strictly increasing transformations of the margins.[5] Under this continuity assumption, the random variable $\boldsymbol{W}=(W_{1},W_{2})$ will have the same copula as the transformed random variable $\boldsymbol{Z}\coloneqq\left(\Phi^{-1}(G_{1}(W_{1})),\Phi^{-1}(G_{2}(W_{2}))\right)$, 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 $\boldsymbol{W}$ 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 $g(\boldsymbol{w})$ (of the original random variable $\boldsymbol{W}$) at a point $\boldsymbol{v}$, find a Gaussian approximation of the pdf

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

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

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

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

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

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

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

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 $K_{\boldsymbol{b}}$ in eq. (2.3). Briefly summarised, the 5×5 covariance matrix of the estimate $\widehat{\boldsymbol{\theta}}_{\boldsymbol{b}}$ will have the form $V_{\boldsymbol{b}}^{-1} W_{\boldsymbol{b}} V_{\boldsymbol{b}}^{-1}$, the presence of the kernel $K_{\boldsymbol{b}}$ means that the matrices $V_{\boldsymbol{b}}$ and $W_{\boldsymbol{b}}$ have rank one in the limit $\boldsymbol{b} \to \boldsymbol{0}^+$, 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 $\boldsymbol{Z}$ 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 estimation[8] 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 $\boldsymbol{W}$ itself is Gaussian. However, the simplified LGC will in general deviate from the normalised LGC-version in section 2.1.2 –

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

[8] Note 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 $\boldsymbol{v} = (v_1, v_2)$, for a strictly stationary univariate time series $\{Y_t\}_{t\in\mathbb{Z}}$ is constructed in the following manner.*

(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

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

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

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

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

Remark 2.2. The requirement $\sum_{h\in\mathbb{Z}} \left|\rho_{\boldsymbol{v}|p}(h)\right| < \infty$ 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 $\boldsymbol{v} \in \mathbb{R}^2$.

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 $t_1, t_2, \ldots, t_n \in \mathbb{Z}$, the vectors $\left(Y_{t_1}, Y_{t_2} \ldots, Y_{t_n}\right)$ and $\left(Y_{-t_1}, Y_{-t_2} \ldots, Y_{-t_n}\right)$ have the same joint distributions.*

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

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

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

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

$$f_{\boldsymbol{v}|p}(\omega) = \overline{f_{\breve{\boldsymbol{v}}|p}(\omega)}. \tag{2.11b}$$

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

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

(d) $f_{\boldsymbol{v}|p}(\omega)$ will in general be complex-valued, but it will always be real valued when the point $\boldsymbol{v}$ 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 $\rho_{\boldsymbol{v}|p}(h)$ 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. $\rho_{\boldsymbol{v}|p}(-h) = \rho_{\breve{\boldsymbol{v}}|p}(h)$, and the definition of time reversibility, see appendices C.1 and C.2 for details. □

Remark 2.3. For general points $\boldsymbol{v} = (v_1, v_2)$, the complex valued result of $f_{\boldsymbol{v}|p}(\omega)$ 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 $f_{\boldsymbol{v}|p}(\omega)$ for $\boldsymbol{v}$ 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 $f_{\boldsymbol{v}|p}(\omega)$ might be interpreted as indicators of e.g. *periodicities at a local level*. This implies that estimates of $f_{\boldsymbol{v}|p}(\omega)$ 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 $\left\{\rho_{\boldsymbol{v}|p}(h)\right\}_{h\in\mathbb{Z}}$ might not be non-negative definite. Caution is thus advised if peaks and troughs of $f_{\boldsymbol{v}|p}(\omega)$ 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

[9] A 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 $\widehat{f}_{\boldsymbol{v}|p}^m(\omega)$ of $f_{\boldsymbol{v}|p}(\omega)$ 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 $\left\{\widehat{z}_t \coloneqq \Phi^{-1}\left(\widehat{G}_n(y_t)\right)\right\}_{t=1}^n$ that corresponds to $\{y_t\}_{t=1}^n$.

(b) Create the lag h pseudo-normalised pairs $\left\{\left(\widehat{z}_{t+h}, \widehat{z}_t\right)\right\}_{t=1}^{n-h}$ for $h = 1, \ldots, m$, and estimate, both for the point $\boldsymbol{v} = (v_1, v_2)$ and its diagonal reflection $\breve{\boldsymbol{v}} = (v_2, v_1)$, the local Gaussian autocorrelations $\left\{\widehat{\rho}_{\boldsymbol{v}|p}(h|\boldsymbol{b}_h)\right\}_{h=1}^m$ and $\left\{\widehat{\rho}_{\breve{\boldsymbol{v}}|p}(h|\boldsymbol{b}_h)\right\}_{h=1}^m$, where the $\{\boldsymbol{b}_h\}_{h=1}^m$ is the bandwidths used during the estimation of the local Gaussian autocorrelation for the different lags.

(c) Adjust eq. (2.11a) *from lemma 2.3(b) with some lag-window function $\lambda_m(h)$ to get the estimate*

$$\widehat{f}_{\boldsymbol{v}|p}^m(\omega) \coloneqq 1 + \sum_{h=1}^m \lambda_m(h) \cdot \widehat{\rho}_{\breve{\boldsymbol{v}}|p}(h|\boldsymbol{b}_h) \cdot e^{+2\pi i\omega h} + \sum_{h=1}^m \lambda_m(h) \cdot \widehat{\rho}_{\boldsymbol{v}|p}(h|\boldsymbol{b}_h) \cdot e^{-2\pi i\omega h}. \quad (2.13)$$

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

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

Moreover, eq. (2.14) *will always hold when the point $\boldsymbol{v}$ 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 logspline technique like the one implemented in Otneim and Tjøstheim (2016).

Remark 2.6. The bandwidths $\boldsymbol{b}_h = (b_{h1}, b_{h2})$ in definition 2.4(b) does not need to be equal for all the lags h when an estimate $\widehat{f}_{\boldsymbol{v}|p}^m(\omega)$ 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 $\boldsymbol{b} = (b_1, b_2)$ such that $b_{hi} \asymp b_i$ for $i = 1, 2$ and for all h (that is to say, $\lim b_{hi}/b_i = 1$).

Remark 2.7. The asymptotic theory for $\widehat{\rho}_{\boldsymbol{v}|p}(h|\boldsymbol{b}_h)$, 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 $\boldsymbol{v}$ for the lag h pairs $\left\{\left(y_{t+h}, y_t\right)\right\}_{t=1}^{n-h}$, i.e. the original observations $\{y_t\}_{t=1}^n$ were used instead of the normalised observations $\left\{z_t \coloneqq \Phi^{-1}\left(G(y_t)\right)\right\}_{t=1}^n$. 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 $\widehat{\rho}_{\boldsymbol{v}|p}(h|\boldsymbol{b}_h)$ 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 $\widehat{f}^m_{\boldsymbol{v}|p}(\omega)$ must in addition consider the size of m relative to both n and the bandwidths $\{\boldsymbol{b}_h\}_{h=1}^m$, 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 $\boldsymbol{v}$, i.e. whether the point $\boldsymbol{v}$ lies at the center or in the periphery of the pseudo-normalised observations $\left\{\left(\widehat{z}_{t+h}, \widehat{z}_t\right)\right\}_{t=1}^{n-h}$. Confer section 3.2 for further details.

Figure 1 shows the effect of the pseudo-normalisation on the `dmbp` example[10] 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.

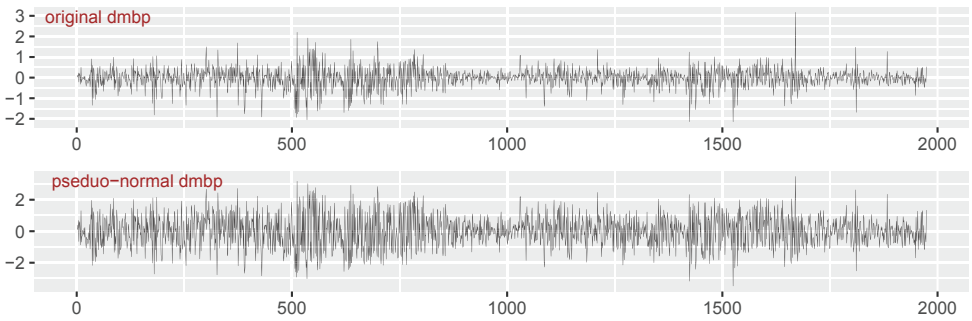

Figure 1: `dmbp`, original version and pseudo-normalised version.

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

This section presents asymptotic results for the cases where $\widehat{f}^m_{\boldsymbol{v}|p}(\omega)$ 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 $100\left[\ln\left(P_t\right) - \ln\left(P_t - 1\right)\right]$, where P_t is the bilateral Deutschemark/British pound rate constructed from the corresponding U.S. dollar rates.'

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 $\widehat{f}^{m}_{\boldsymbol{v}|p}(\omega)$ 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,*

$$\boldsymbol{Y}_{h:t} \coloneqq \left[Y_{t+h}, Y_t\right]', \qquad \boldsymbol{Y}_{\bar{m}:t} \coloneqq \left[Y_{t+m}, \ldots, Y_t\right]', \tag{2.15}$$

and let $g_h(\boldsymbol{y}_h)$ and $g_{\bar{m}}(\boldsymbol{y}_{\bar{m}})$ 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 $\widetilde{\eta}_h(\boldsymbol{y}_{\bar{m}}) : \mathbb{R}^{m+1} \to \mathbb{R}^1$ is the *obvious extension*[11] of a bivariate function $\eta_h(\boldsymbol{y}_h) : \mathbb{R}^2 \to \mathbb{R}^1$, then

$$\mathrm{E}[\eta_h(\boldsymbol{Y}_{h:t})] = \mathrm{E}[\widetilde{\eta}_h(\boldsymbol{Y}_{\bar{m}:t})]\,, \qquad \text{for } h \in \{1, \ldots, m\}\,. \tag{2.16}$$

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 $X^n_{hq:i}$ in item (g) in essence are the result that occurs when the product of the kernel function $K_{h:b}(\boldsymbol{y}_h - \boldsymbol{v})$ and the score function of the local Gaussian approximation $\psi(\boldsymbol{y}_h)$ is evaluated in $\boldsymbol{y}_h = \boldsymbol{Y}_{h:t}$.

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

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

[11] The 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

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

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

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

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

$$g_h(\boldsymbol{y}_h) = g_h(\boldsymbol{v}) + \boldsymbol{\mathfrak{g}}_h(\boldsymbol{v})' \left[\boldsymbol{y}_h - \boldsymbol{v}\right] + \boldsymbol{\mathfrak{R}}_h(\boldsymbol{y}_h)' \left[\boldsymbol{y}_h - \boldsymbol{v}\right], \tag{2.18}$$

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

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

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

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

$$\boldsymbol{0} < \boldsymbol{b}_0 \coloneqq \inf_{h\in\mathbb{Z}} \boldsymbol{b}_{h0}, \tag{2.19}$$

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

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

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 $h \to \infty$, i.e. the bivariate density functions $g_h(\boldsymbol{y}_h)$ will for large lags h approach the product of the marginal densities, and the situation will thus 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 $\boldsymbol{b}_0$ 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 $K_{\boldsymbol{b}}$ and the score function of the bivariate Gaussian distribution, see lemma C.6 for details.

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

The score function in eq. (2.4), i.e. $\boldsymbol{u}(\boldsymbol{w};\boldsymbol{\theta}) \coloneqq \frac{\partial}{\partial \boldsymbol{\theta}} \log\left(\psi(\boldsymbol{w};\boldsymbol{\theta})\right)$, plays a central role in the local density-estimation approach of Hjort and Jones (1996), and it 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 $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}} - \boldsymbol{\theta}_{\boldsymbol{v}}$ does implicitly require that $\boldsymbol{u}(\boldsymbol{v};\boldsymbol{\theta}_{\boldsymbol{v}}) \neq \mathbf{0}$ in order for the corresponding asymptotic covariance matrix to be well defined. The investigation of $\left(\widehat{f}^{\,m}_{\boldsymbol{v}|p}(\omega) - f_{\boldsymbol{v}|p}(\omega)\right)$ 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 $\psi_p(\boldsymbol{w};\boldsymbol{\theta}_p)$ that is considered, and that $\boldsymbol{u}_p(\boldsymbol{w};\boldsymbol{\theta}_p)$ here represent the corresponding score function.

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

(a) $\boldsymbol{u}_p(\boldsymbol{v};\boldsymbol{\theta}_{p|\boldsymbol{v}}(h)) \neq \mathbf{0}$ for all finite h.

(b) $\lim \boldsymbol{u}_p(\boldsymbol{v};\boldsymbol{\theta}_{p|\boldsymbol{v}}(h)) \neq \mathbf{0}$.

Remark 2.13. It is, for a given time series Y_t and a given point $\boldsymbol{v}$, possible to inspect the p equations in $\boldsymbol{u}_p(\boldsymbol{w};\boldsymbol{\theta}_p) = \mathbf{0}$ 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 $\boldsymbol{\theta}^{\dagger}_{p|\boldsymbol{v}}$ which solve $\boldsymbol{u}_p(\boldsymbol{v};\boldsymbol{\theta}_p) = 0$, and then observe that assumption 2.2(a) will fail if $\boldsymbol{\theta}^{\dagger}_{p|\boldsymbol{v}} \in \left\{\boldsymbol{\theta}_{p|\boldsymbol{v}}(h)\right\}$. 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 $h \to \infty$. Together with the assumption of normalised marginals, this implies that the limit of $\boldsymbol{\theta}_{p|\boldsymbol{v}}(h)$ always becomes $[\mu_1, \mu_2, \sigma_1, \sigma_2, \rho]' = [0, 0, 1, 1, 0]'$, which means that assumption 2.2(b) will fail for any point $\boldsymbol{v}$ that solves $\boldsymbol{u}_p(\boldsymbol{v}; [0, 0, 1, 1, 0]') = 0$.

Remark 2.14. The one parameter local Gaussian approximation $\psi_1(\boldsymbol{w};\boldsymbol{\theta}_1)$ is less flexible than the five parameter approximation $\psi_5(\boldsymbol{w};\boldsymbol{\theta}_5)$, 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 $\boldsymbol{v}$, see the discussion in appendix C.6.2 for further details.

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

For simplicity, the present analysis will use the $\boldsymbol{b} = (b_1, b_2)$ introduced in remark 2.6, see page 10, i.e. it will be assumed that the individual bandwidths $\boldsymbol{b}_h$ for the different lags h approach zero at the same rate – and that it for the asymptotic investigation thus can be assumed that the same bandwidth is used for all the lags. For the present case, where the lag h pairs are of the form $\left(Y_{t+h}, Y_t\right)$ 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 $m \coloneqq m_n \to \infty$ be a sequence of integers denoting the number of lags to include, and let $\boldsymbol{b} \coloneqq \boldsymbol{b}_n \to \mathbf{0}^+$ be the bandwidths used when estimating the local Gaussian correlations for the lags $h = 1, \ldots, m$ (based on n observations). Let b_1 and b_2 refer to the two components of $\boldsymbol{b}$, and let α, ν and a be as introduced in assumption 2.1(b). Let $s \coloneqq s_n \to \infty$ be a sequence of integers such that $s = o\left(\sqrt{nb_1b_2/m}\right)$, and let τ be a positive constant. The

following requirements must be satisfied for these entities.[12]

(a) $\log n/n(b_1 b_2)^5 \longrightarrow 0$, (only required for the case $p = 5$).
(b) $nb_1 b_2/m \longrightarrow \infty$.
(c) $m^\delta(b_1 \vee b_2) \longrightarrow 0$, where $\delta = 2 \vee \frac{\nu(a+1)}{\nu(a-1)-2}$.
(d) $\sqrt{nm/b_1 b_2} \cdot s^\tau \cdot \alpha(s - m + 1) \longrightarrow \infty$.
(e) $m = o\big((nb_1 b_2)^{\tau/(2+5\tau)-\lambda}\big)$, 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 $\widehat{\rho}_{\boldsymbol{v}|5}(h)$.

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 $\widehat{f}^m_{\boldsymbol{v}|p}(\omega)$

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

$$\sqrt{n(b_1 b_2)^{(p+1)/2}/m} \cdot \Big(\widehat{f}^m_{\boldsymbol{v}|p}(\omega) - f_{\boldsymbol{v}|p}(\omega)\Big) \xrightarrow{d} \mathrm{N}\big(\mathbf{0}, \sigma^2_{\boldsymbol{v}|p}(\omega)\big)\,, \tag{2.20}$$

where the formula

$$\sigma^2_{\boldsymbol{v}|p}(\omega) = 4 \lim_{m\to\infty} \frac{1}{m} \sum_{h=1}^{m} \lambda^2_m(h) \cdot \cos^2(2\pi\omega h) \cdot \tilde{\sigma}^2_{\boldsymbol{v}|p}(h) \tag{2.21}$$

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

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 $f_{\boldsymbol{v}|p}(\omega)$ is a real valued function for all points $\boldsymbol{v}$ 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 $\widehat{f}^m_{\boldsymbol{v}|p}(\omega)$.*

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

[12] Notational 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 $\widehat{f}^{m}_{\boldsymbol{v}|p}(\omega)$ complex-valued is given in appendix A.2, where it can be seen that $\sqrt{n(b_1 b_2)^{(p+1)/2}/m} \cdot \left(\widehat{f}^{m}_{\boldsymbol{v}|p}(\omega) - f_{\boldsymbol{v}|p}(\omega)\right)$ 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 $\widehat{f}^{m}_{\boldsymbol{v}|p}(\omega)$ 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`-data[13] 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 $\widehat{f}^{m}_{\boldsymbol{v}|p}(\omega)$ 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 $\widehat{f}^{m}_{\boldsymbol{v}|p}(\omega)$ 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 $\widehat{f}^{m}_{\boldsymbol{v}|p}(\omega)$, 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 $\widehat{f}^{m}_{\boldsymbol{v}|p}(\omega)$ 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

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

densities, but the estimated parameters $\widehat{\rho}_{\boldsymbol{v}|1}(h)$ might not give a good indicator of the local dependency-structure of the targeted distribution. The 5-parameter estimates $\widehat{\rho}_{\boldsymbol{v}|5}(h)$ have thus been selected for all of the plots, with the sole exception of fig. 14 where two plots based on $\widehat{\rho}_{\boldsymbol{v}|1}(h)$ 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 $f^m_{\boldsymbol{v}|p}(\omega)$.

3.1 Setting the input parameters

Several input parameters must be selected before an estimate of $\widehat{f}^m_{\boldsymbol{v}|p}(\omega)$ 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. $\boldsymbol{v}$, 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. $\boldsymbol{b} = (b_1, b_2)$, the bandwidth-vector to be used when computing the local Gaussian autocorrelations. Most of the plots shown in this section have used $\boldsymbol{b} = (.5, .5)$, with the exception of fig. 6, where plots based on $\boldsymbol{b} = (.75, .75)$ and $\boldsymbol{b} = (1, 1)$ have been included for comparison.[15]
4. m, the truncation level, i.e. the number of lags to include in the estimate of $\widehat{f}^m_{\boldsymbol{v}|p}(\omega)$. 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.

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

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

[14] The pointwise confidence band gives for each frequency ω a confidence interval for the value of $f^m_{\boldsymbol{v}|p}(\omega)$.

[15] It 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.

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

First of all, note that the combination of point $\boldsymbol{v}$ and bandwidth $\boldsymbol{b}$ influences how many of the h-lagged pairs that effectively contributes to the computation of $\widehat{\rho}_{\boldsymbol{v}|5}(h)$. 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 $\boldsymbol{v}$, 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 $\boldsymbol{v}$.

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 $\boldsymbol{v}$ and bandwidth $\boldsymbol{b}$) 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 $\left(\widehat{z}_{t+h}, \widehat{z}_t\right)$ to occur within a lag h bandwidth-square (centered at a diagonal point $\boldsymbol{v}$), it is necessary that both $\widehat{z}_{t+h}$ and $\widehat{z}_t$ 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 or 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`.

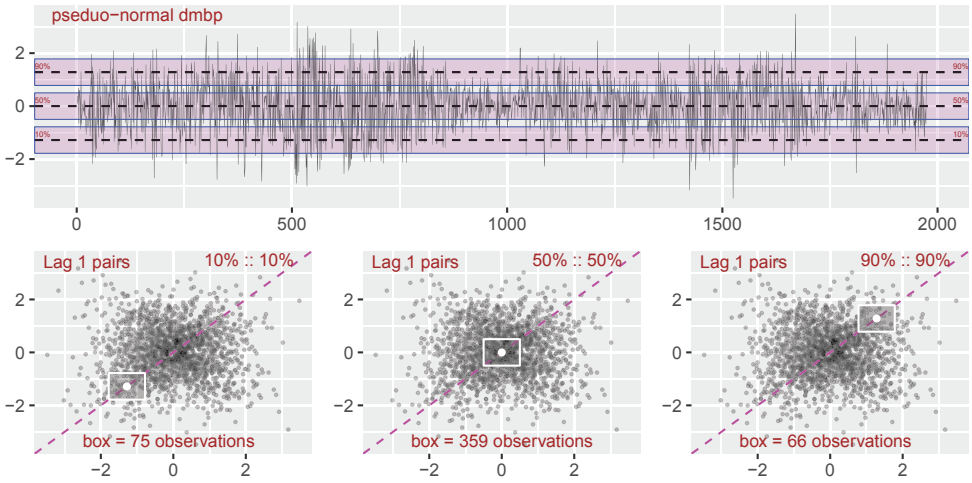

Figure 2: `dmbp` (pseudo-normalised version), *levels* and *bandwidth-bands* (top) and *lag 1 bandwidth-squares* (bottom).

be an asymmetry between the lower and upper tails of the `dmbp`-data. Furthermore, as the bias-variance properties of the estimates $\widehat{\rho}_{\boldsymbol{v}|5}(h)$ depends on the number of points that effectively contributes during the computation, it is clear that the variance will increase for points $\boldsymbol{v}$ that lie farther out in the tails. The selection of which tail-points to investigate must thus take into account the number of available observations for the lags to be included.

Remark 3.11. An important detail with regard to the estimation of $f_{\boldsymbol{v}|5}(\omega)$ is the selection of the truncation level m, since that value (in addition to the value of the bandwidth $\boldsymbol{b}$) influences the bias-variance properties of the estimate $\widehat{f}^{\,m}_{\boldsymbol{v}|5}(\omega)$. 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 $f_{\boldsymbol{v}|5}(\omega)$.

Remark 3.12. If the truncation level m is too large, the interconnection between m and $\boldsymbol{b}$ could create a situation (for points $\boldsymbol{v}$ in the periphery of the data) where the number of lag h pseudo-normalised observations used to estimate $\widehat{\rho}_{\boldsymbol{v}|5}(h)$ 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, $\boldsymbol{v}$ and $\boldsymbol{b}$ 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 $f_{\boldsymbol{v}|5}(\omega)$ 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 $\widehat{\rho}_{\boldsymbol{v}|5}(h)$ varies for the three points of interest (when $\boldsymbol{b} = (0.5, 0.5)$). Red dotted lines shows the truncation level $m = 10$ (to be used later on), in order to emphasise which estimates of $\widehat{\rho}_{\boldsymbol{v}|5}(h)$ that will contribute to the estimation of $\widehat{f}^{m}_{\boldsymbol{v}|5}(\omega)$. This plot shows that there is a clear distinction between the center and the two tails. The $\widehat{\rho}_{\boldsymbol{v}|5}(h)$ tends to fluctuate around 0 at the center, which implies that the corresponding estimated spectral density $\widehat{f}^{m}_{\boldsymbol{v}|5}(\omega)$ most likely will be rather flat and close to 1. For the two tails, it seems natural to assumme that some long-range dependency must be present, and one might also suspect that there is an asymmetry between the two tails.

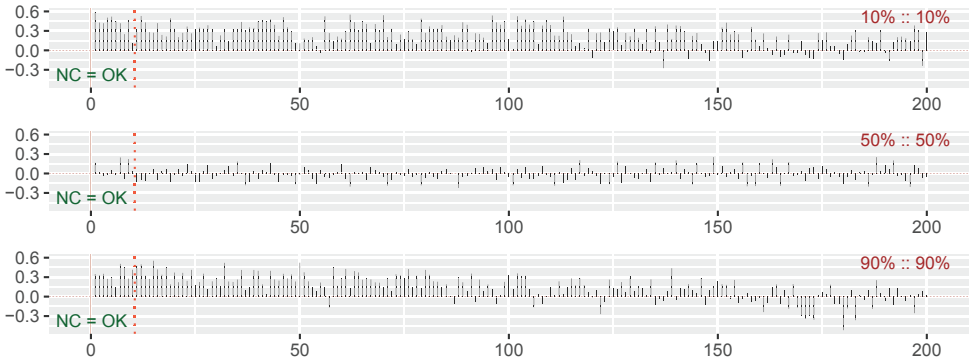

Figure 3: dmbp-data, $\widehat{\rho}_{\boldsymbol{v}|5}(h)$ for $h = 1, \ldots, 200$ (for the three points of interest).

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 $\widehat{\rho}_{\boldsymbol{v}|5}(h)$ automatically implies that the m-truncated estimated spectral density $\widehat{f}^{m}_{\boldsymbol{v}|5}(\omega)$ 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 $\widehat{f}^{m}_{\boldsymbol{v}|5}(\omega)$ for the dmbp-data are presented.

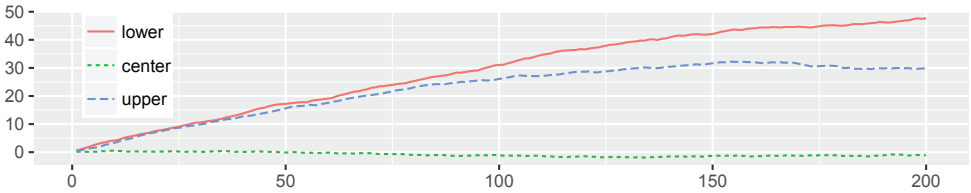

Figure 4: dmbp-data, cumulative sum of local Gaussian auto-correlations.

3.3 Some simulations

This section will estimate the local Gaussian spectral densities $f_{\boldsymbol{v}|5}(\omega)$ for simulated data. It is known from lemma 2.3(a) that $f_{\boldsymbol{v}|5}(\omega)$ 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 $\widehat{f}^m_{\boldsymbol{v}|5}(\omega)$ 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 $f_{\boldsymbol{v}|5}(\omega)$ and $f(\omega)$, see the definition below.

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

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

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

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 $\mathrm{N}(0, 1)$. The computations are based on the bandwidth $\boldsymbol{b} = (0.5, 0.5)$, and the points (on the diagonal) corresponds to the 0.1, 0.5 and 0.9 quantiles of the standard normal distribution. The top left panel shows the pseudo-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 $f^m_{\boldsymbol{v}|5}(\omega)$ for the three points under investigation (blue part). Information about the truncation level and the points are printed at the top of each plot.

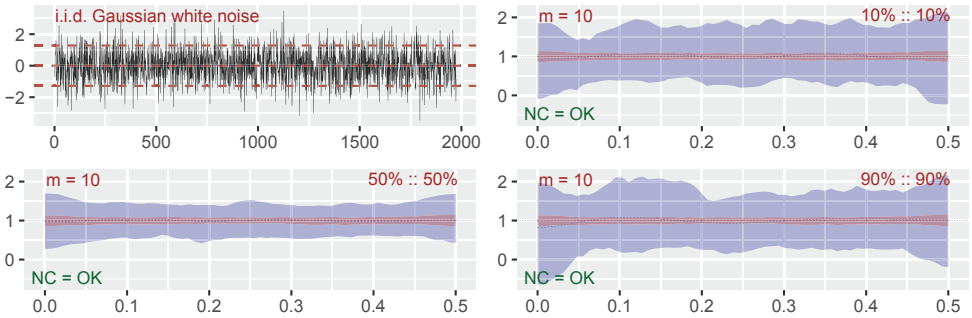

Figure 5: i.i.d. Gaussian white noise.

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 $f^m_{\boldsymbol{v}|5}(\omega)$ (the m-truncations of the true values), which in this case in fact coincides with $f(\omega)$ and $f_{\boldsymbol{v}|5}(\omega)$, 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 $f_{\boldsymbol{v}|5}(\omega)$ 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 $f^m_{\boldsymbol{v}|5}(\omega)$. It might be necessary to apply a (much) higher truncation level before $f^m(\omega)$ and $f^m_{\boldsymbol{v}|5}(\omega)$ gives decent approximations of the true values $f(\omega)$ and $f_{\boldsymbol{v}|5}(\omega)$. It thus seems preferable to estimate $\widehat{f}^m_{\boldsymbol{v}|5}(\omega)$ 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 $\boldsymbol{v}$ (given a suitable bandwidth $\boldsymbol{b}$). These artificial time series will in 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 $\boldsymbol{b} = (.5, .5)$ (the bandwidth used in all the other examples), whereas the two plots at the bottom shows the situation for the bandwidths $\boldsymbol{b} = (.75, .75)$ and $\boldsymbol{b} = (1, 1)$, respectively at

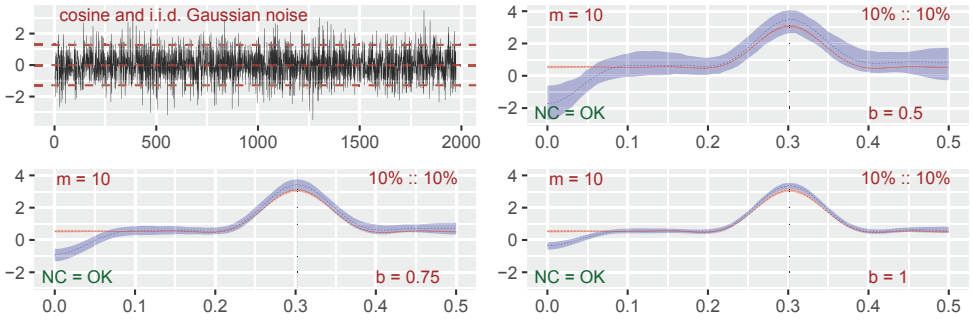

Figure 6: Single cosine and i.i.d. white noise, same point, bandwidths based on 0.5, 0.75 and 1.

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 $\boldsymbol{b} = (.5, .5)$ 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 $K_{\boldsymbol{b}}$ that estimates $\rho_{\boldsymbol{v}|5}(h)$, 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 $\{\rho_{\boldsymbol{v}|5}(h)\}_{h\in\mathbb{Z}}$, the collection of local Gaussian autocorrelations, may not be a non-negative definite function. In fact, based on the observation that the estimates of $\widehat{f}^m_{\boldsymbol{v}|5}(\omega)$ 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 $\widehat{f}^m_{\boldsymbol{v}|5}(\omega)$ 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 $\widehat{f}^m_{\boldsymbol{v}|5}(\omega)$, 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 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 $(L_1, \ldots, L_r)$ at the y-axis, a collection of amplitudes $(A_1, \ldots, A_r)$, a collection of frequencies $(\alpha_1, \ldots, \alpha_r)$ and a collection of phase-adjustments $(\phi_1, \ldots, \phi_r)$. Finally, assign a probability p_i to each

$i = 1, \ldots, r$, 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 $(A'_1, \ldots, A'_r)$. 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,

$$C_i(t) = L_i + A_i(t) \cdot \cos\left(2\pi\alpha_i t + \phi_i\right), \qquad i = 1, \ldots, r, \tag{3.2}$$

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

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

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 sample[18] 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 $\boldsymbol{v}$ (given a suitable bandwidth $\boldsymbol{b}$). 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.

[18] The 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.

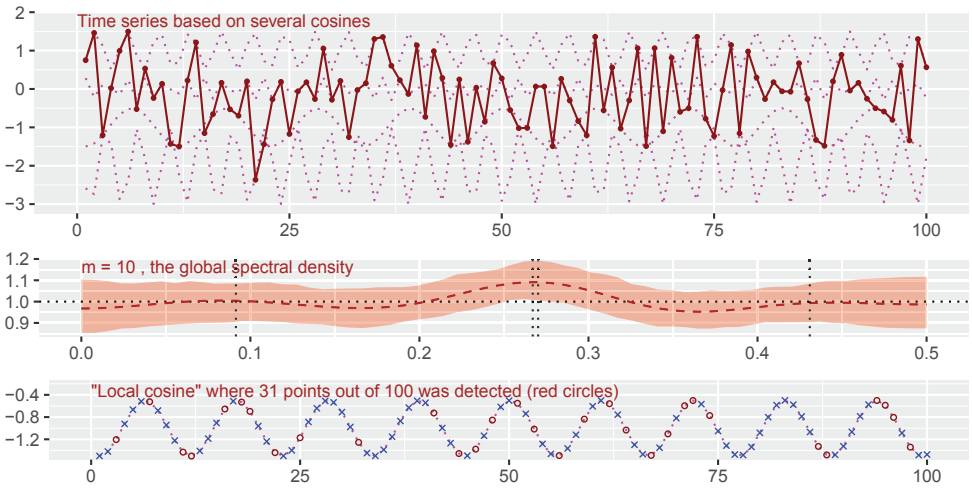

Figure 7: Top: Short excerpt from artifical example based on *hidden trigonometric components*. Center: Estimated (truncated) global spectral density (*hidden frequencies* indicated with vertical lines). Bottom: Local cosine showing the detected points at the local level centered at -1.

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 $\boldsymbol{b}$) 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 $\boldsymbol{v}$) is computed from the sets of bivariate points $\left(Y_{t+h}, Y_t\right)$. 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 $\widehat{f}^m_{\boldsymbol{v}|5}(\omega)$ is presented in fig. 8, as usual based on 100 independent samples of length 1974, a bandwidth $\boldsymbol{b} = (0.5, 0.5)$ 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, $\widehat{f}^m_{\boldsymbol{v}|5}(\omega)$ 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 $\widehat{f}^{m}_{\boldsymbol{v}|5}(\omega)$ 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.

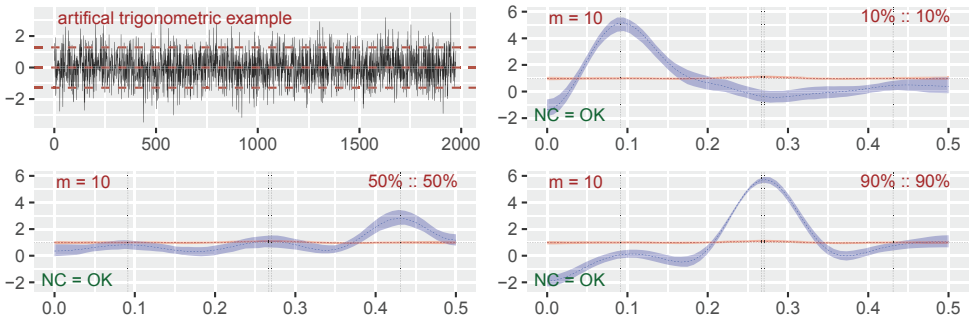

Figure 8: Artifical example, *hidden trigonometric components*.

3.3.3 Beware of global structures

It is important to keep in mind that a comparison of the local Gaussian spectral density $f_{\boldsymbol{v}|5}(\omega)$ 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 $f_{\boldsymbol{v}|5}(\omega)$ 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-

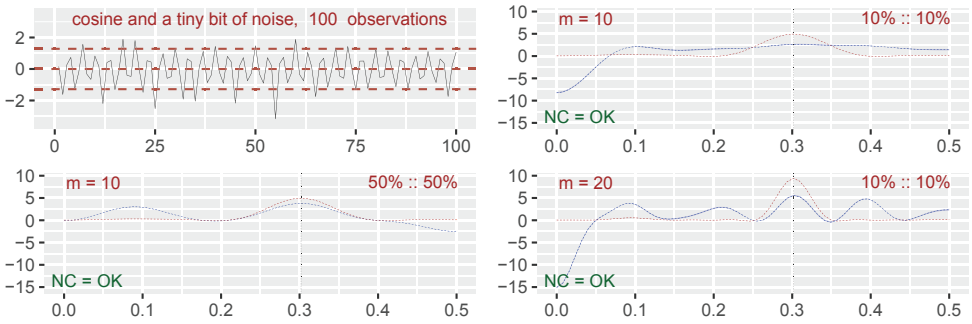

Figure 9: Pseudo-normalised single cosine and a tiny bit of noise.

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), were it was given as

$$\epsilon_t = s_t e_t, \qquad e_t \sim \mathrm{N}(0,1)\,, \tag{3.4a}$$

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

$$\alpha_0 > 0, \delta \geq 0, \quad \alpha_i \geq 0, i = 1\ldots, p, \quad -1 < \gamma_i < 1, i = 1\ldots, p, \quad \beta_j \geq 0, j = 1\ldots, q, \tag{3.4c}$$

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 $\boldsymbol{b} = (0.5, 0.5)$. 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 $f^m_{v|5}(\omega)$ 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.

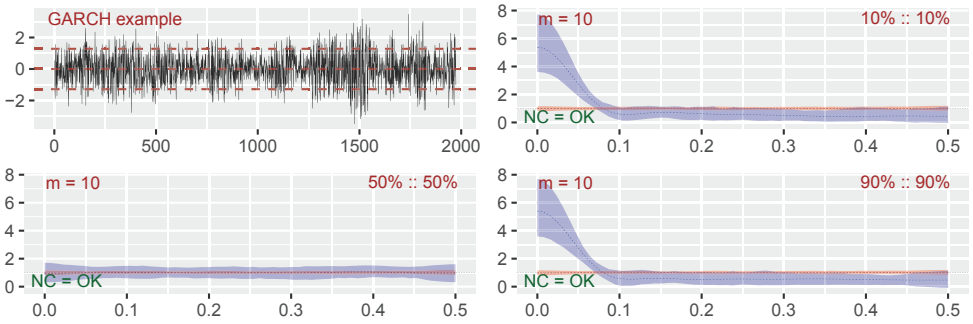

Figure 10: GARCH-type model, based on dmbp.

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 $f^m_{\boldsymbol{v}|5}(\omega)$ 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 $\widehat{f}^m_{\boldsymbol{v}|5}(0) = 1 + 2 \cdot \sum_{h=1}^{m} \lambda_m(h) \cdot \widehat{\rho}_{\boldsymbol{v}|5}(h)$, 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 $\widehat{\rho}_{\boldsymbol{v}|5}(h)$ for the dmbp-data, upon which the parameters of the apARCH-model was based. Furthermore, the fact that $f^m_{\boldsymbol{v}|5}(\omega)$ seems to be very close to 1 at the center indicates that the estimated local Gaussian autocorrelations $\widehat{\rho}_{\boldsymbol{v}|5}(h)$ 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 $\widehat{\rho}_{\boldsymbol{v}|5}(h)$ 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 $\widehat{f}^m_{\boldsymbol{v}|5}(\omega)$ for a given point $\boldsymbol{v}$ at a given frequency ω requires a selection of a bandwidth $\boldsymbol{b}$ 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 $\boldsymbol{b} = (0.5, 0.5)$, 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.

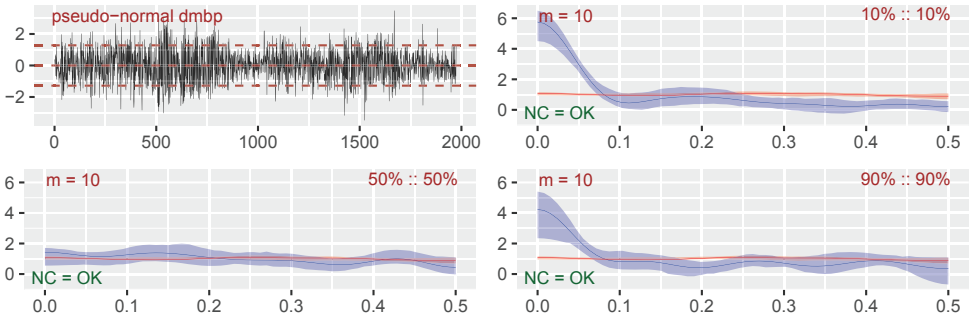

Figure 11: dmbp-data, bootstrapped based confidence intervals.

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 $\boldsymbol{b} = (0.5, 0.5)$.

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.

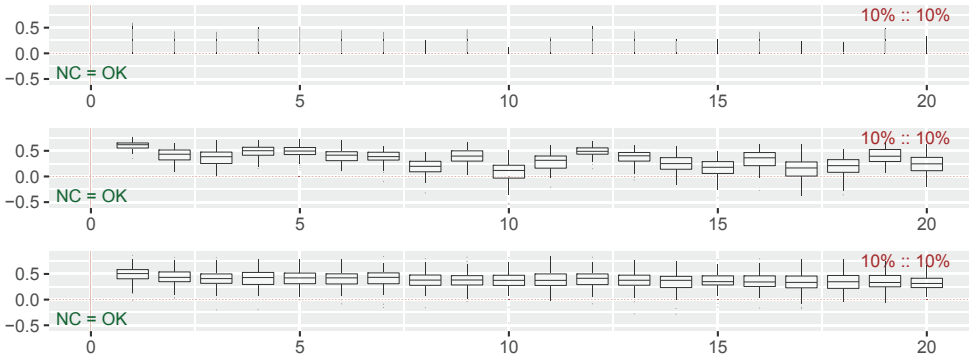

Figure 12: Local Gaussian auto-correlations related to the pseudo-normalised dmbp-data. Top: Original data. Center: Bootstrapped results (100 replicated using the blocklength 100). Bottom: apARCH(2,3) fitted to dmbp (100 simulations).

3.6 Exploration for off-diagonal points

The preceding examples all considered points $\boldsymbol{v} = (v_1, v_2)$ on the diagonal, i.e. $v_1 = v_2$, but mathematically it is as such not a problem to estimate $f_{\boldsymbol{v}|5}(\omega)$ 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 $\widehat{f}^m_{\boldsymbol{v}|5}(\omega)$ 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 $\boldsymbol{v} = (-1.28, 1.28)$, i.e. the first coordinate corresponds to the 10% quantile and the second coordinate corresponds to the 90% quantile.

Since the estimates $\widehat{f}^m_{\boldsymbol{v}|5}(\omega)$ 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 $\widehat{f}^{\,m}_{\boldsymbol{v}|5}(\omega) = \widehat{c}^{\,m}_{\boldsymbol{v}|5}(\omega) - i\widehat{q}^{\,m}_{\boldsymbol{v}|5}(\omega)$, 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 $\widehat{c}^{\,m}_{\boldsymbol{v}|5}(\omega)$ and $\widehat{q}^{\,m}_{\boldsymbol{v}|5}(\omega)$ will respectively be referred to as estimates of the *auto-cospectrum* and the *auto-quadrature spectrum* of $\widehat{f}^{\,m}_{\boldsymbol{v}|5}(\omega)$.

The bias-variance properties of $\widehat{\rho}_{\boldsymbol{v}|5}(h)$ 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 $\left(\widehat{z}_{t+h}, \widehat{z}_t\right)$ 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 $\widehat{\rho}_{\boldsymbol{v}|5}(h)$ (for $h = 1, \ldots, 10$) 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 $\widehat{c}^{\,m}_{\boldsymbol{v}|5}(\omega)$ are given at the top and the estimated auto-quadrature spectra $\widehat{q}^{\,m}_{\boldsymbol{v}|5}(\omega)$ 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 $\widehat{c}^{\,m}_{\boldsymbol{v}|5}(\omega)$, and only a dotted horizontal line at $y = 0$ has been added to the plot that investigates $\widehat{q}^{\,m}_{\boldsymbol{v}|5}(\omega)$.

Remark 3.20. For this particular point $\boldsymbol{v}$ and this particular truncation-level m, the estimated auto-cospectra $\widehat{c}^{\,m}_{\boldsymbol{v}|5}(\omega)$ 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 $\widehat{q}^{\,m}_{\boldsymbol{v}|5}(\omega)$ 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.

[19] The 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 $\boldsymbol{b} = (0.5, 0.5)$.

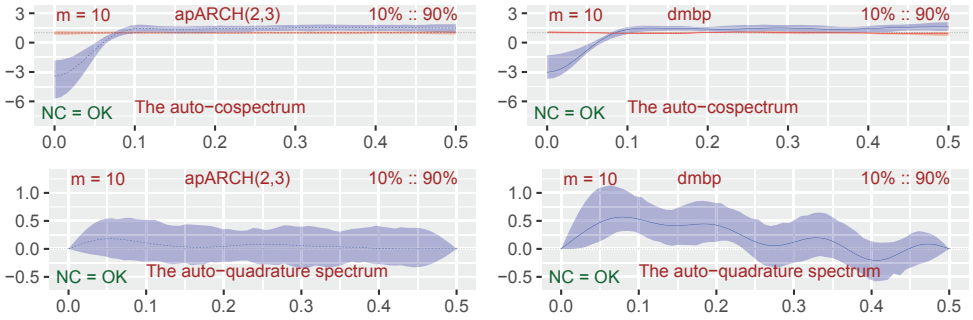

Figure 13: An off-diagonal point, comparison apARCH$(2,3)$-model (left) and `dmbp` (right), auto-cospetrcum (top) vs. auto-quadrature spectrum (bottom).

Remark 3.21. An interpretation of $\widehat{c}^{\,m}_{\boldsymbol{v}|5}(\omega)$ and $\widehat{q}^{\,m}_{\boldsymbol{v}|5}(\omega)$ 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 $\widehat{q}^{\,m}_{\boldsymbol{v}|5}(\omega)$ 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 $\widehat{c}^{\,m}_{\boldsymbol{v}|5}(\omega)$ can reveal the presence of an initial long-range sequence of same-sign estimated local Gaussian autocorrelations $\widehat{\rho}_{\boldsymbol{v}|5}(h)$. Furthermore, from the definitions it follows that $\widehat{c}^{\,m}_{\boldsymbol{v}|5}(\omega)$ and $\widehat{q}^{\,m}_{\boldsymbol{v}|5}(\omega)$ 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 $\widehat{q}^{\,m}_{\boldsymbol{v}|5}(\omega)$ 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 $\widehat{q}^{\,m}_{\boldsymbol{v}|5}(\omega)$-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 $\boldsymbol{v}$, truncation level m and bandwidth $\boldsymbol{b}$ used in this example.

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

Remark 3.24. This example shows that $\widehat{f}^{\,m}_{\boldsymbol{v}|5}(\omega)$ might provide some insight also when the point $\boldsymbol{v}$ lies off the diagonal. In a practical setting, it would of course be natural/necessary to investigate several different off-diagonal points $\boldsymbol{v}$ for different combinations of bandwidths $\boldsymbol{b}$ 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 $\widehat{\rho}_{\boldsymbol{v}|5}(h)$ of $\rho_{\boldsymbol{v}|5}(h)$ have been used in the computations. The reason for this is that $\widehat{\rho}_{\boldsymbol{v}|1}(h)$ in general might not carry

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

The inadequacy of $\widehat{f}^m_{\boldsymbol{v}|1}(\omega)$ is highlighted in fig. 14 where it has been compared against $\widehat{f}^m_{\boldsymbol{v}|5}(\omega)$ 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 $\widehat{\rho}_{\boldsymbol{v}|1}(h)$ was used instead of $\widehat{\rho}_{\boldsymbol{v}|5}(h)$. 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 $\widehat{f}^m_{\boldsymbol{v}|1}(\omega)$ does not detect a difference that is clear-cut when $\widehat{f}^m_{\boldsymbol{v}|5}(\omega)$ is employed. This makes the $\widehat{f}^m_{\boldsymbol{v}|1}(\omega)$ less favourable as an exploratory tool.[20]

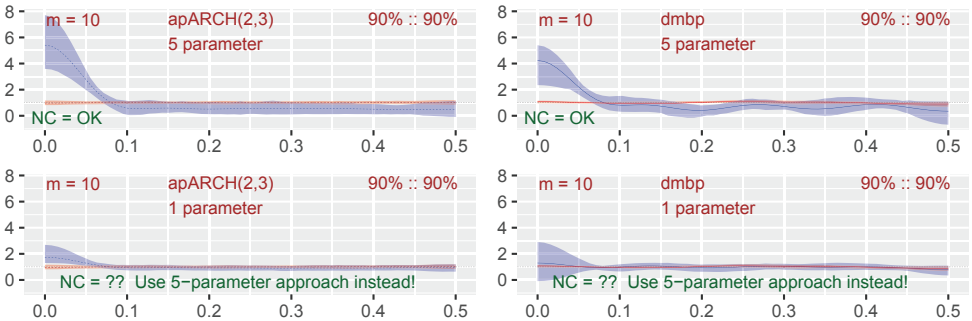

Figure 14: apARCH$(2, 3)$-model (left) and `dmbp` (right), 5 parameter (top) vs. 1 parameter (bottom).

The 'trumpet shaped' pointwise confidence band observed in the lower right panel seems to be a common feature when the 1 parameter approach $\widehat{f}^m_{\boldsymbol{v}|1}(\omega)$ 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 $\widehat{f}^m_{\boldsymbol{v}|5}(\omega)$ 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 $\widehat{f}^m_{\boldsymbol{v}|1}(\omega)$ might not be up to the task when the structure of interest deviates from the Gaussian assumption.

[20] Note that the algorithm used to find $\widehat{\rho}_{\boldsymbol{v}|1}(h)$ 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 $f_{\boldsymbol{v}|5}(\omega)$ 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 $\boldsymbol{v}$ and the parameters m and $\boldsymbol{b}$ 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 $\boldsymbol{v}$ and a bandwidth $\boldsymbol{b}$ has a large impact on the number of pseudo-normalised lag h pairs from $\left\{\left(\widehat{z}_{t+h}, \widehat{z}_t\right)\right\}_{t=1}^{n-h}$ that efficiently contributes to the computation of the estimates $\widehat{\rho}_{\boldsymbol{v}|5}(h)$.

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

In section 3, the bandwidth $\boldsymbol{b} = (.5, .5)$ 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 $f^m_{\boldsymbol{v}|5}(\omega)$ (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 $\boldsymbol{b} = (.5, .5)$ 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 $h = 1, \ldots, m$, and it could also be a problem that the same bandwidth is used for all the points $\boldsymbol{v}$, 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. $\rho_{\boldsymbol{v}|5}(h)$ and $\rho_{\boldsymbol{v}|1}(h)$, 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 $\boldsymbol{v}$ 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 $h = 1, \ldots, m$. In particular, if a selection-algorithm for $\boldsymbol{b}$ could be based on a cross-validation technique, then it would for practical purposes probably be preferable to first compute estimates of $f^m_{\boldsymbol{v}|5}(\omega)$ 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 $\rho_{\boldsymbol{v}|1}(h)$. This option did however result in a tremendous increase in the computational time, since the estimation of $f^m_{\boldsymbol{v}|1}(\omega)$ requires the estimation of m different local Gaussian auto-correlations $\rho_{\boldsymbol{v}|1}(h)$.[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_ν-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 $\widehat{f}(\omega)$. 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 $\widehat{\rho}(h)$ 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 $\widehat{\rho}_{\boldsymbol{v}|5}(h)$ are affected both by the position of the point $\boldsymbol{v}$ and the selected bandwidth $\boldsymbol{b}$. In particular, the kernel function involved in the estimation of $\rho_{\boldsymbol{v}|5}(h)$ implies that the variance will depend on the number of pseudo-normalised lag h pairs $\left(\widehat{z}_{t+h}, \widehat{z}_t\right)$ 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 $\boldsymbol{v}$ with the smallest number of contributing pseudo-normalised pairs should be considered, i.e.

[21] Tests were performed to see if it might be possible to only use the bandwidth-algorithm for the case $h = 1$, and then let the higher lags inherit the estimated bandwidth – but it turned out that that 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 $\widehat{\rho}_{\boldsymbol{v}|5}(h)$ 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 $\boldsymbol{v}$ 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 $\widehat{\rho}_{\boldsymbol{v}|5}(h)$ 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 $\widehat{f}_{\boldsymbol{v}|5}(h)$. 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 $G \coloneqq \sum_{h=-\infty}^{\infty} |h| R(h)$ 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 $\widehat{f}^{\,m}_{\boldsymbol{v}|5}(\omega)$ 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

[22] It 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 $\widehat{\rho}_{\boldsymbol{v}|5}(h)$ instead. A potential problem with this approach is that the result could depend upon the points $\boldsymbol{v}$ that are investigated. But still, if nothing else, a visual inspection (like the one given in fig. 3) of the estimated values $\widehat{\rho}_{\boldsymbol{v}|5}(h)$ 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 $\widehat{f}^{\,m}_{\boldsymbol{v}|5}(\omega)$ 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 $\widehat{f}^{\,m}_{\boldsymbol{v}|5}(\omega)$ of $f_{\boldsymbol{v}|5}(\omega)$ was based on a weighting function $\lambda_{_m}(h)$ that worked upon the estimated values $\widehat{\rho}_{\boldsymbol{v}|5}(h)$, 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 $\widehat{f}^{\,m}_{\boldsymbol{v}_i|5}(\omega)$ corresponding to different points $\{\boldsymbol{v}_i\}_{i=1}^{\nu}$ will be jointly asymptotically normal and pairwise asymptotically independent (when $m \to \infty$ and $\boldsymbol{b} \to \mathbf{0}^+$ as $n \to \infty$). This enables an alternative smoothing strategy, where an estimate $\widehat{f}^{\,m}_{\boldsymbol{v}|5}(\omega)$ for a given point $\boldsymbol{v}$ could be based on a weighting of the values of $\widehat{f}^{\,m}_{\boldsymbol{v}_i|5}(\omega)$ in a grid of points surrounding $\boldsymbol{v}$.

This alternative approach shares some superficial similarities with the one used when the ordinary global spectrum $f(\omega)$ is computed based on the periodogram, see e.g. Brockwell and Davis (1986) for details. However, the efficiency of the periodogram-approach in the estimation of $f(\omega)$ is due to the *Fast Fourier Transform*, which implies that the periodogram can be computed directly from the observations without the need for an explicit computation of all of the 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* $f_{\boldsymbol{v}|p}(\omega)$ 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 $\rho_{\boldsymbol{v}|p}(h)$.

Estimates $\widehat{f}^{m}_{\boldsymbol{v}|p}(\omega)$ 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 examples[23] 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.

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

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

This appendix presents the asymptotic properties of $\widehat{f}^m_{\boldsymbol{v}|p}(\omega)$, 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 $\widehat{f}^m_{\boldsymbol{v}|p}(\omega)$ 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 $f_{\boldsymbol{v}|p}(\omega)$ is a real-valued function when $\boldsymbol{v}$ lies on the diagonal was proved in lemma 2.3(d). The expression for $\widehat{f}^m_{\boldsymbol{v}|p}(\omega)$ from lemma 2.5 can by the help of vectors be written as

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

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

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

$$\widehat{\boldsymbol{P}}_{\boldsymbol{v}|m|\boldsymbol{b}|p} \coloneqq \left[\widehat{\rho}_{\boldsymbol{v}|p}(1|\boldsymbol{b}_1), \ldots, \widehat{\rho}_{\boldsymbol{v}|p}(m|\boldsymbol{b}_m)\right]'. \tag{A.2b}$$

Since $\widehat{\rho}_{\boldsymbol{v}|p}(h|\boldsymbol{b}_h)$ is one of the p estimated parameters $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|p}(h|\boldsymbol{b}_h)$ from the local Gaussian approximation (of the lag h pairs) at the point $\boldsymbol{v}$,[24] it is clear that it is possible to write $\widehat{\rho}_{\boldsymbol{v}|p}(h|\boldsymbol{b}_h) = \boldsymbol{e}'_p \cdot \widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|p}(h|\boldsymbol{b}_h)$, where $\boldsymbol{e}'_p$ is the unit vector that picks out $\widehat{\rho}_{\boldsymbol{v}|p}(h|\boldsymbol{b}_h)$ from $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|p}(h|\boldsymbol{b}_h)$. The vectors $\left\{\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|p}(h|\boldsymbol{b}_h)\right\}_{h=1}^m$ can be stacked on top of each other to give a joint parameter vector $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}|p}$, and it follows that the vector $\widehat{\boldsymbol{P}}_{\boldsymbol{v}|m|\boldsymbol{b}|p}$ can be expressed as $\widehat{\boldsymbol{P}}_{\boldsymbol{v}|m|\boldsymbol{b}|p} = \boldsymbol{E}'_{m|p} \cdot \widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}|p}$, where $\boldsymbol{E}'_{m|p}$ is the matrix that picks out the relevant components from $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}|p}$. ($\boldsymbol{E}'_{m|p}$ 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 $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}|p}$ will give an asymptotic normality result for $\widehat{f}^m_{\boldsymbol{v}|p}(\omega)$. In particular, if a suitable scaling factor[25] $c_{n|m|\boldsymbol{b}|p}$ gives an mp-variate asymptotic normality result for $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}|p}$,

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

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

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

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

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

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

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

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

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

from which a simple calculation gives

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

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

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

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

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

$$\sigma^2_{c:\boldsymbol{v}|p}(\omega) = \lim_{m\to\infty} \frac{1}{m} \sum_{h=1}^{m} \lambda^2_m(h) \cdot \cos^2(2\pi\omega h) \cdot \left\{ \tilde{\sigma}^2_{\boldsymbol{v}|p}(h) + \tilde{\sigma}^2_{\breve{\boldsymbol{v}}|p}(h) \right\} \tag{A.9a}$$

$$\sigma^2_{q:\boldsymbol{v}|p}(\omega) = \lim_{m\to\infty} \frac{1}{m} \sum_{h=1}^{m} \lambda^2_m(h) \cdot \sin^2(2\pi\omega h) \cdot \left\{ \tilde{\sigma}^2_{\boldsymbol{v}|p}(h) + \tilde{\sigma}^2_{\breve{\boldsymbol{v}}|p}(h) \right\}, \tag{A.9b}$$

with $\tilde{\sigma}^2_{\boldsymbol{v}|p}(h)$ *and* $\tilde{\sigma}^2_{\breve{\boldsymbol{v}}|p}(h)$ *related to respectively* $\widehat{\rho}_{\boldsymbol{v}|p}(h|\boldsymbol{b}_h)$ *and* $\widehat{\rho}_{\breve{\boldsymbol{v}}|p}(h|\boldsymbol{b}_h)$ *as given in theorem 2.7.*

The component $\widehat{q}^{\,m}_{\boldsymbol{v}|p}(\omega)$ *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*

$$\sqrt{n(b_1 b_2)^{(p+1)/2}/m} \cdot \left(\widehat{f}^{\,m}_{\boldsymbol{v}|p}(\omega) - f_{\boldsymbol{v}|p}(\omega) \right) \xrightarrow{d} \mathrm{N}\big(0, \sigma^2_{c:\boldsymbol{v}|p}(\omega)\big), \tag{A.10}$$

Proof. The case $\omega \in \frac{1}{2}\cdot\mathbb{Z}$ can be proved by the exact same argument that was used in the proof of theorem 2.7, whereas the general case requires a bivariate extension of that proof. In particular, when the proof of theorem 2.7 is used on $\widehat{c}^{\,m}_{\boldsymbol{v}|p}(\omega)$ and $\widehat{q}^{\,m}_{\boldsymbol{v}|p}(\omega)$, it follows that they can be written as

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

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

where $\boldsymbol{\Lambda}'_{c|m}(\omega)$ and $\boldsymbol{\Lambda}'_{q|m}(\omega)$ are the coefficient vectors containing respectively the cosines and sines, where $\widehat{\boldsymbol{P}}_{\boldsymbol{v}|m|\boldsymbol{b}|p}$ and $\widehat{\boldsymbol{P}}_{\breve{\boldsymbol{v}}|m|\boldsymbol{b}|p}$ contains the estimated correlations corresponding to $\boldsymbol{v}$ and $\breve{\boldsymbol{v}}$ for the lags under consideration, and where the length $2m$ vectors $\boldsymbol{\Lambda}'_{c|\overline{\underline{m}}}(\omega)$, $\boldsymbol{\Lambda}'_{q|\overline{\underline{m}}}(\omega)$ and $\widehat{\boldsymbol{P}}_{\boldsymbol{v}|\overline{\underline{m}}|\boldsymbol{b}|p}$ 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 $\widehat{\boldsymbol{P}}_{\boldsymbol{v}|\overline{\underline{m}}|\boldsymbol{b}|p} = \left(\boldsymbol{E}'_{m|p} \oplus \boldsymbol{E}'_{m|p}\right) \cdot \widehat{\boldsymbol{\Theta}}_{\overline{\underline{m}}|\boldsymbol{b}|p}(\boldsymbol{v}, \breve{\boldsymbol{v}})$, where $\widehat{\boldsymbol{\Theta}}_{\overline{\underline{m}}|\boldsymbol{b}|p}(\boldsymbol{v}, \breve{\boldsymbol{v}})$ is the full set of estimated parameters from the local Gaussian approximations at $\boldsymbol{v}$ and $\breve{\boldsymbol{v}}$ for the lags under consideration,[26] and where $\left(\boldsymbol{E}'_{m|p} \oplus \boldsymbol{E}'_{m|p}\right)$ is the matrix that picks out the relevant autocorrelations.

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

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

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

$$\sqrt{n(b_1 b_2)^{(p+1)/2}} \cdot \left(\widehat{\boldsymbol{\Theta}}_{\overline{\underline{m}}|\boldsymbol{b}|p}(\boldsymbol{v}, \breve{\boldsymbol{v}}) - \boldsymbol{\Theta}_{\overline{\underline{m}}|p}(\boldsymbol{v}, \breve{\boldsymbol{v}}) \right) \xrightarrow{d} \mathrm{N}\big(\boldsymbol{0}, \Sigma_{\boldsymbol{v}|\overline{m}|p} \oplus \Sigma_{\breve{\boldsymbol{v}}|\overline{m}|p}\big), \tag{A.13}$$

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 $\sigma^2_{q:\boldsymbol{v}|p}(\omega)$ 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. □

[26] The vector $\widehat{\boldsymbol{\Theta}}_{\overline{\underline{m}}|\boldsymbol{b}|p}(\boldsymbol{v}, \breve{\boldsymbol{v}})$ can be expressed as a combination of $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}|p}$ and $\widehat{\boldsymbol{\theta}}_{\breve{\boldsymbol{v}}|\overline{m}|\boldsymbol{b}|p}$, where $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}|p}$ is the parameter vector from the proof of theorem 2.7.

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

This section will investigate the asymptotic properties of the parameter vector $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}|p}$, 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 $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|p}(h|\boldsymbol{b}_h)$, i.e. the Klimko-Nelson penalty function approach will be used to derive the desired result.

Appendix B.1 explains the Klimko-Nelson approach and shows how a local penalty function for the present case can be constructed based on the local penalty function encountered in (Tjøstheim and Hufthammer, 2013). Appendix B.2 verifies the fourth of the requirements needed for the Klimko-Nelson approach, and the asymptotic results for $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}|p}$ 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 $\boldsymbol{v}$, $\boldsymbol{b}$ 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* $\rho_{\boldsymbol{v}|5}$ as a new *local measure of dependence* at a point $\boldsymbol{v}$, and then used the approach formalised in Klimko and Nelson (1978), to investigate the asymptotic properties of $\widehat{\rho}_{\boldsymbol{v}|5}$. (The arguments in (Tjøstheim and Hufthammer, 2013) holds for the simpler case $p = 1$ too.) The *local Gaussian spectral density* $f_{\boldsymbol{v}|p}(\omega)$ is based on the local Gaussian autocorrelations $\rho_{\boldsymbol{v}|p}(h)$, and the asymptotic properties of the estimates $\widehat{f}^{m}_{\boldsymbol{v}|p}(\omega)$ are thus closely connected to the asymptotic properties of $\widehat{\rho}_{\boldsymbol{v}|p}(h)$.

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

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

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

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

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

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

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

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

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

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

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

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

then

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

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

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

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

a minimiser of which should satisfy the vector equation

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

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

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

For a sample of size n from $\{\boldsymbol{Y}_{h:t}\}_{t\in\mathbb{Z}}$, the following M-estimator[28] will be used, which (due to the ergodicity implied by assumption 2.1(a)) will converge towards the penalty function $q_{h:\boldsymbol{b}}$,

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

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

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

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

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

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

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

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

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

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

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

where

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

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

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

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

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

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

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

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

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

B.1.3 A new penalty function

The proof of theorem 2.7 requires an asymptotic result for the parameter vector $\widehat{\boldsymbol{\theta}}_{n|\overline{m}|\boldsymbol{b}|p}$, which was obtained by combining m parameter vectors corresponding to the bivariate lag h pairs $\left(Y_{t+h}, Y_t\right)$ for $h = 1, \ldots, m$. This section will show (for the case $p = 5$) how a penalty function for $\widehat{\boldsymbol{\theta}}_{n|\overline{m}|\boldsymbol{b}|p}$ can be constructed based on the bivariate penalty functions $Q_{h:n}$ defined in appendix B.1.2. The indices n, $\boldsymbol{b}$ and p will for notational simplicity be suppressed from the notation, and only $\boldsymbol{\theta}_{\overline{m}}$ will henceforth be used.

An analysis akin to the one in Theorem 1 of (Tjøstheim and Hufthammer, 2013) will be performed in this section, i.e. the asymptotic situation will be investigated for the simple case where the truncation m and the bandwidth $\boldsymbol{b}$ both are fixed as $n \to \infty$. The proof that the

new penalty function satisfies the four requirements items *(A1)* to *(A4)* of theorem B.1 can then be based upon corresponding components of the proof of Theorem 1 from (Tjøstheim and Hufthammer, 2013).

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

With regard to the construction of the new penalty function, the main observation of interest is that the $Q_{h:n}(\boldsymbol{\theta}_h)$ from appendix B.1.2 was defined for bivariate time series $\{\boldsymbol{Y}_{h:t}\}_{t\in\mathbb{Z}}$, whereas the new penalty function will be defined for the $(m+1)$-variate time series $\{\boldsymbol{Y}_{\overline{m}:t}\}_{t\in\mathbb{Z}}$. The first step is to extend the penalty functions $Q_{h:n}$, $h = 1, \ldots, m$ from expression based on $\boldsymbol{Y}_{h:t}$ to expressions based on $\boldsymbol{Y}_{\overline{m}:t}$, but this is trivial since the bivariate functions occurring in the definition of $Q_{h:n}(\boldsymbol{\theta}_h)$ can be extended in a natural manner to $(m+1)$-variate functions, as mentioned in remark 2.9, which gives the desired functions $\widetilde{Q}_{h:n}(\boldsymbol{\theta}_h)$.

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

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

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

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

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

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

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

where $\boldsymbol{\theta}_{\overline{m}:\boldsymbol{b}}$ represents the vector obtained by stacking on top of each other the m individual population parameters $\boldsymbol{\theta}_{h:\boldsymbol{b}}$, where $\boldsymbol{\nabla}_{\overline{m}} \coloneqq [\boldsymbol{\nabla}'_1, \ldots, \boldsymbol{\nabla}'_m]'$, and where the matrices $V_{\overline{m}|\boldsymbol{b}:n}$ and $T_{\overline{m}|\boldsymbol{b}:n}$ corresponds to the matrices $V_{h:\boldsymbol{b}:n}$ and $T_{h:\boldsymbol{b}:n}$ 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 $\boldsymbol{\nabla}_{\overline{m}}$ is defined relative to $\boldsymbol{\nabla}_h$, and how $Q_{\overline{m}:n}$ is defined relative to $Q_{h:n}$, it is clear that $\boldsymbol{\nabla}_{\overline{m}} Q_{\overline{m}:n}(\boldsymbol{\theta}_{\overline{m}:\boldsymbol{b}})$ is the vector obtained by stacking the m vectors $\boldsymbol{\nabla}_h Q_{h:n}(\boldsymbol{\theta}_{h:\boldsymbol{b}})$ on top of each other.
2. The operator $\boldsymbol{\nabla}_{\overline{m}} \boldsymbol{\nabla}_{\overline{m}}'$ can be viewed as an $m \times m$ block-matrix, consisting of the 5×5 matrices $\boldsymbol{\nabla}_j \boldsymbol{\nabla}_k'$, $j, k = 1, \ldots, m$. Due to the definition of $Q_{\overline{m}:n}$, it is clear that the only operators $\boldsymbol{\nabla}_j \boldsymbol{\nabla}_k'$ that will return a nonzero result are those having $j = k$.
3. The preceding observation implies that $V_{\overline{m}|\boldsymbol{b}:n} = \bigoplus_{h=1}^{m} V_{h:\boldsymbol{b}:n}$, i.e. $V_{\overline{m}|\boldsymbol{b}:n}$ is the direct sum of the matrices $V_{h:\boldsymbol{b}:n}$ (the block diagonal matrix where the diagonal blocks equals $V_{h:\boldsymbol{b}:n}$, and all other blocks are zero, cf. e.g. Horn and Johnson (2012, p.30) for further details).
4. The same observation implies that $T_{\overline{m}|\boldsymbol{b}:n} = \bigoplus_{h=1}^{m} T_{h:\boldsymbol{b}:n}$

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

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

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

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

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

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

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

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

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

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

Proof. As done in the proof of Theorem 1 in (Tjøstheim and Hufthammer, 2013), the idea is to first prove asymptotic normality of each individual component of $\boldsymbol{\nabla}_{\overline{m}}Q_{\overline{m}:n}(\boldsymbol{\theta}_{\overline{m}:\boldsymbol{b}})$ 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 $\boldsymbol{\nabla}_{\overline{m}}Q_{\overline{m}:n}(\boldsymbol{\theta}_{\overline{m}:\boldsymbol{b}})$ will be the joint distribution of these limiting components, and finally a simple observation based on moment-generating functions tells us that this limiting joint distribution is asymptotically normal.

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

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

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

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

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

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 $\mathrm{E}\big[|X_{hi:t}|^\nu\big] < \infty$. Finally, since $\boldsymbol{\theta}_{h:\boldsymbol{b}}$ is the population value parameter that minimise eq. (B.5), it follows that $\mathrm{E}[X_{hi:t}] = 0$. These observations show that $X_{hi:t}$ satisfies the requirements needed in order to apply Theorem 2.20(i) and Theorem 2.21(i) from (Fan and Yao, 2003, p. 74-75), i.e. for $S_{hi|n} \coloneqq \sum_{t=1}^{n} X_{hi:t}$, Theorem 2.20(i) gives the asymptotic result

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

with 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

B.2 The A4-requirement in the general case

The verification of the three first requirements of the Klimko-Nelson approach does work as before when '$m \to \infty$ and $\boldsymbol{b} \to \boldsymbol{0}^+$ when $n \to \infty$', whereas the asymptotic normality in the fourth requirement demands a more detailed investigation. Appendix B.2.1 will introduce some new building blocks to be used in the investigation of the asymptotic properties, which will be developed in appendices B.2.2 and B.2.3. Some technical details that only depend upon the kernel function and the score functions have been collected in appendix C.4.

B.2.1 The final building blocks

The bivariate processes $\boldsymbol{Y}_{h:t}$ from definition 2.6 will now be used to construct new random variables, that culminates in a random variable $\mathfrak{Q}_{\overline{m}}^{n}$ which has the same limiting distribution[29] $\sqrt{b_1 b_2}\boldsymbol{\nabla}_{\overline{m}} Q_{\overline{m}:n}(\boldsymbol{\theta}_{\overline{m}:\boldsymbol{b}})$. Looking upon eq. (B.8), it is clear that everything depends upon the three functions $\psi(\boldsymbol{y}_h;\boldsymbol{\theta}_h)$, $\boldsymbol{u}_h(\boldsymbol{y}_h;\boldsymbol{\theta}_h)$ and $K_{h:\boldsymbol{b}}(\boldsymbol{y}_h - \boldsymbol{v})$. 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 $\psi(\boldsymbol{y}_h;\boldsymbol{\theta}_h)$ the local Gaussian density used when approximating $g_h(\boldsymbol{y}_h)$ at the point $\boldsymbol{v} = (v_1, v_2)$, define for all $h \in \mathbb{N}$ and $q \in \{1,\dots,p\}$*

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

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

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

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

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

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

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

The following simple observations will be useful later on.

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

(a) $\sup_{hq}\big|u_{hq:\boldsymbol{b}}(\boldsymbol{v})\big| < \infty$ and $\sup_{hq}\big|u_{hq}(\boldsymbol{v})\big| < \infty$.

(b) When L is large enough, $u_{hq:\boldsymbol{b}}(\boldsymbol{v})^{\leq L} = u_{hq:\boldsymbol{b}}(\boldsymbol{v})$ and $u_{hq}(\boldsymbol{v})^{\leq L} = u_{hq}(\boldsymbol{v})$.

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

The bivariate kernel to be used in the present approach will be the same as the one used in Tjøstheim and Hufthammer (2013), i.e. it will be the product kernel based on two standard

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

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

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 $K(\boldsymbol{w})$, that satisfies*

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

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

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

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

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

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

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

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

Definition B.10. *With $g_h(\boldsymbol{y}_h)$, $u_{hq:\boldsymbol{b}}(\boldsymbol{w})$ and $K(\boldsymbol{w})$ as given in definitions 2.6, B.7 and B.9, let $\mathfrak{U}_{h:\boldsymbol{b}} \coloneqq \left[u_{h1:\boldsymbol{b}}(\boldsymbol{v}), \ldots, u_{hp:\boldsymbol{b}}(\boldsymbol{v})\right]'$, and define the following matrices.*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The last batch of random variables can now be introduced.

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

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

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

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

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

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

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

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

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

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 $\boldsymbol{b}$ goes to zero when n and m are 'large enough'. The proof will be presented in a step by step manner, that builds upon the asymptotic behaviour of $\mathrm{E}\big[X_{hq:i}^{n} \cdot X_{jr:k}^{n}\big]$. The computation of this expectation will (depending on the indices h, i, j and k) either require a bivariate, trivariate or tetravariate integral.

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

Table 1: Factors deciding bivariate, trivariate or tetravariate.

Table 1 lists the combinations that must be taken into account when computing $\mathrm{E}\big[X_{hq:i}^{n} \cdot X_{jr:k}^{n}\big]$, i.e. the presence of $\boldsymbol{v}$ and $\boldsymbol{b}$ and the dependence on Y_t in the kernel functions – and it is evident from this table that the amount of overlap in the indexing set $\{i, h+i, k, j+k\}$ will decide if the resulting integral turns out to be bi-, tri- or tetravariate. Note that eq. (2.13) 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 $K(\boldsymbol{w})$ or the score-function components $u_{hq:\boldsymbol{b}}(\boldsymbol{w})$ had bounded support, since the finiteness requirements of assumption 2.1(g) then would follow directly from lemma C.6, and the proof of lemma B.15 would be rather trivial. However, in the present analysis, $K(\boldsymbol{w})$ and $u_{hq:\boldsymbol{b}}(\boldsymbol{w})$ both have $\mathbb{R}^2$ as their support, which implies that extra care must be taken when working with the densities under consideration.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Since assumption 2.1(g) ensures that the sum of the tree 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.

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

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

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

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

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

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

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

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

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

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

Proof. The connection between expectations and L^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

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

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

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

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

B.2.3 The asymptotic results – final part

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

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

Lemma B.18. *When Y_t satisfies assumption 2.1, when n, m and $\boldsymbol{b}$ are as specified in assumption 2.3, and when $u_{hq:\boldsymbol{b}}(\boldsymbol{w})$ and $K(\boldsymbol{w})$ are as given in definitions B.7 and B.9 – then the random variables $Z^n_{\bar{m}:t}(\boldsymbol{a})$ from definition B.13 satisfies*

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

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

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

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

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

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

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

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

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

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

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

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

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

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;

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\operatorname{Var}(Z_{\overline{m}:t}^{n}(\boldsymbol{a})) = \sum_{h=1}^{m}\sum_{j=1}^{m}\sum_{q=1}^{p}\sum_{r=1}^{p} a_{hq} a_{jr} \operatorname{Cov}\big(X_{hq:t}^{n}, X_{jr:t}^{n}\big) \tag{B.51a}$$

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

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

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

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

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

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

The main theorem can now be stated, i.e. this result can be used to verify the fourth requirement of the Klimko-Nelson approach for the penalty function $Q_{\overline{m}:n}(\boldsymbol{\theta}_{\overline{m}:\boldsymbol{b}})$, from which it follows an asymptotic normality result for $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}|p}$, that finally gives the asymptotic normality result of $\widehat{f}_{\boldsymbol{v}|p}^{m}(\omega)$. (Confer remark B.3 for an interpretation of the m that occurs in the limiting distributions.)

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

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

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

$$\mathcal{Q}^n_{\bar{m}}(\boldsymbol{a}) = \sum_{h=1}^m \sum_{q=1}^p a_{hq} \mathcal{Q}^n_{hq} = \sum_{h=1}^m \sum_{q=1}^p a_{hq} \left[\sum_{t=1}^n Z^n_{hq:t}\right] = \sum_{t=1}^n \left[\sum_{h=1}^m \sum_{q=1}^p a_{hq} Z^n_{hq:t}\right] = \sum_{t=1}^n Z^n_{\bar{m}:t}(\boldsymbol{a}). \tag{B.53a}$$

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

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

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

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

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

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

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

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

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

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

such that

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

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

$$\operatorname{Var}(n^{-1/2}\,\wp_{\bar{m}}^{n}(\boldsymbol{a})) = \frac{1}{n}\,\mathrm{E}[\wp_{\bar{m}}^{n}(\boldsymbol{a})\cdot\wp_{\bar{m}}^{n}(\boldsymbol{a})] = \frac{1}{n}\sum_{p=1}^{3}\sum_{q=1}^{3}\mathrm{E}[S_n^{(p)}\cdot S_n^{(q)}]\,. \tag{B.58}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Assumption 2.3(f), i.e. $m = o(s)$, ensures that there (asymptotically) will be no overlap between these 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 $\mathfrak{s}_\ell^2 \coloneqq \sum_{j=1}^{\ell} \mathrm{Var}\left(\eta_j\right)$ is needed. From assumption 2.1(a) and corollary B.19(b), it follows that

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

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

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

This holds trivially if the sets occurring in the indicator functions, i.e. $\left\{\left|\eta_j\right| \geq \epsilon\sqrt{\mathfrak{s}_\ell^2}\right\}$, becomes empty when n is large enough. It is thus of interest to see if an upper bound for $\left|\eta_j\right|$ 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 $X^n_{hq:t}$, $Z^n_{hq:t}$ and η_j, see eqs. (B.30a), (B.32a) and (B.56a), it is clear that an upper bound for $\left|\eta_j\right|$ might be deduced from,

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

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

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

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

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

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

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

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

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

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

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

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

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

[32] Truncation 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

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

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

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

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

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

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 $\boldsymbol{b}$. One part of the 'asymptotic problem' is the interpretation of an infinite-variate Gaussian distribution, but the main problem is the occurrence of the kernel function $K(\boldsymbol{w})$, which in the limit gives a degenerate Gaussian distribution in theorem B.20(b) (when $p > 1$). This degeneracy in itself would not have been any issue if the target of interest had been the asymptotic behaviour of $n^{-1/2}\,\mathfrak{Q}_{\bar{m}}^n$, 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 $\widehat{\boldsymbol{\theta}}_{\bar{m}:n}$, see appendix B.3 for details.

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

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

i.e. asymptotically mp-variate normal.

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

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

The final details needed for the investigation of the asymptotic properties of $\widehat{f}_{\boldsymbol{v}|p}^{m}(\omega)$ 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 $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|\bar{m}|\boldsymbol{b}|p}$ converges towards the true parameter vector $\boldsymbol{\theta}_{\boldsymbol{v}|\bar{m}|p}$ in the following manner.*

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

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

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

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

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

with $\Sigma_{\boldsymbol{v}|\overline{m}|p} \coloneqq V^{-1}_{\boldsymbol{v}|\overline{m}|p} W_{\boldsymbol{v}|\overline{m}|p} V^{-1}_{\boldsymbol{v}|\overline{m}|p}$, where the $mp \times mp$ matrices $V_{\boldsymbol{v}|\overline{m}|p}$ and $W_{\boldsymbol{v}|\overline{m}|p}$ can be represented as

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

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

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

For the general situation, when $m \to \infty$ and $\boldsymbol{b} \to \mathbf{0}^+$ when $n \to \infty$, it is necessary with an additional scaling in order to get a covariance matrix with finite entries. Obviously, a factor $\sqrt{b_1 b_2}$ must be included in order to balance the effect of the kernel function $K_{h:\boldsymbol{b}}$ – and for the $p = 1$ case this is sufficient since the matrices $V_{\boldsymbol{v}|h|p}$ and $W_{\boldsymbol{v}|h|p}$ reduces to nonzero scalars.

For the $p = 5$ case, the limiting matrices of $V_{\boldsymbol{v}|h|p}$ and $W_{\boldsymbol{v}|h|p}$ 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 $p = 5$. □

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

The previous analysis was restricted to the case where one point was used throughout, which is sufficient for the investigation of the asymptotic properties of the m-truncated estimates $\widehat{f}^{\,m}_{\boldsymbol{v}|p}(\omega)$ for a point $\boldsymbol{v}$ that lies upon the diagonal (see theorem 2.7) or for general points $\boldsymbol{v} \in \mathbb{R}^2$ when the time series under investigation is time reversible (see theorem 2.8).

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

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

$$\sqrt{n(b_1 b_2)^{(p+1)/2}} \cdot \left(\widehat{\boldsymbol{\Theta}}_{\underline{\overline{m}}|\boldsymbol{b}|p}(\boldsymbol{v}, \breve{\boldsymbol{v}}) - \boldsymbol{\Theta}_{\underline{\overline{m}}|p}(\boldsymbol{v}, \breve{\boldsymbol{v}})\right) \xrightarrow{d} \mathrm{N}\left(\boldsymbol{0}, \begin{bmatrix} \Sigma_{\boldsymbol{v}|\overline{m}|p} & 0 \\ 0 & \Sigma_{\breve{\boldsymbol{v}}|\overline{m}|p} \end{bmatrix}\right), \tag{B.77}$$

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

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

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

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

The requirement that $\boldsymbol{v}$ lies off the diagonal together with the requirement that none of the bivariate densities $g_h(\boldsymbol{y}_h)$ possess diagonal symmetry implies that different approximating local Gaussian densities occurs for the different points and different lags, so it can be assumed that there is no common parameters in $\boldsymbol{\theta}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}|p}$ and $\boldsymbol{\theta}_{\breve{\boldsymbol{v}}|\overline{m}|\boldsymbol{b}|p}$. This implies that the arguments used to verify the three first requirements of theorem B.1 for the penalty function $Q_{\overline{m}:n|p}$ (see lemmas B.3 to B.5), also will work upon the combined penalty function $Q_{\underline{\overline{m}}:n|p}$, and it will in particular be the case that the Hessian matrix $V_{\underline{\overline{m}}|\boldsymbol{b}:n|p}$ occurring in lemma B.4 can be written as the direct sum of the matrices that corresponds to $Q_{\overline{m}:n|p}\big(\boldsymbol{\theta}_{\boldsymbol{v}|\overline{m}|\boldsymbol{b}|p}\big)$ and $Q_{\overline{m}:n|p}\big(\boldsymbol{\theta}_{\breve{\boldsymbol{v}}|\overline{m}|\boldsymbol{b}|p}\big)$, i.e. $V_{\underline{\overline{m}}|\boldsymbol{b}|p}(\boldsymbol{v}, \breve{\boldsymbol{v}}) = V_{\overline{m}|\boldsymbol{b}:n|p}(\boldsymbol{v}) \oplus V_{\overline{m}|\boldsymbol{b}:n|p}(\breve{\boldsymbol{v}})$, 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 $Q_{\underline{\overline{m}}:n|p}$ requires some minor modifications of the constructions that was encountered in appendix B.2.1. Both $X^n_{hq:t}(\boldsymbol{v})$ and $X^n_{hq:t}(\breve{\boldsymbol{v}})$ (for $h = 1, \ldots, m$ and $q = 1, \ldots, p$) are needed, and the final random variable will include both $\boldsymbol{v}$ and $\breve{\boldsymbol{v}}$ versions of the variables $Z^n_{hq:t}$, $\mathcal{Q}^n_{hq}$, $Z^n_{\overline{m}:t}(\boldsymbol{a})$, $\boldsymbol{Z}^n_{\overline{m}:t}$, $\mathcal{Q}^n_{\overline{m}}(\boldsymbol{a})$ and $\boldsymbol{\mathcal{Q}}^n_{\overline{m}}$.

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

The statement for $Z^n_{\overline{m}:t}(\boldsymbol{a})$ given in corollary B.17(b) extends trivially to the present case, since the asymptotic behaviour are unaffected by the adjustment that a sum of length m is

replaced by two sums of length m. The statement in lemma B.18 remains the same too, but some minor adjustments are needed in the proof: First of all, from the definition of $Z^n_{\underline{\overline{m}}:t}(\boldsymbol{a}_1, \boldsymbol{a}_2; \boldsymbol{v}, \breve{\boldsymbol{v}})$, it follows that

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}_i|\overline{m}|\boldsymbol{b}|p}$ corresponding to different points $\{\boldsymbol{v}_i\}_{i=1}^{\nu}$ will be jointly asymptotically normal and pairwise asymptotically independent. The asymptotically independent property are inherited by the corresponding estimated local Gaussian spectral densities $\widehat{f}^{m}_{\boldsymbol{v}_i|p}(\omega)$, and this enables an alternative smoothing strategy for the estimated local Gaussian spectral densities at a given point $\boldsymbol{v}$, 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 $\rho_{\boldsymbol{v}|p}(h)$ 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 $\boldsymbol{b}$, whereas appendix C.4 shows that the assumptions on the kernel function $K(\boldsymbol{w})$ and the score functions $u_{hq:\boldsymbol{b}}(\boldsymbol{w})$ implies that some integrals are finite (which implies that assumption 2.1(g) will be trivially satisfied if the bivariate densities $g_h(\boldsymbol{y}_h)$ are finite). Appendix C.5 contains a few basic definitions/comments related to 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 $\widehat{f}^{m}_{\boldsymbol{v}|5}(\omega)$ should be used instead of $\widehat{f}^{m}_{\boldsymbol{v}|1}(\omega)$.

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

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

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

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

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

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

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

gives the following property[33] for the probability density functions,

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

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

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 $f_{\boldsymbol{v}|p}(\omega)$ 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*

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

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

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

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

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

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

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.*

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

(a) There exists integers s that makes items (e) and (f) of assumption 2.3 compatible.
(b) There exists integers s and constants $c \coloneqq c_{_n} \to \infty$, such that

$$c \cdot s = o\Big(\sqrt{nb_{_1}b_{_2}/m}\Big)\,, \qquad \sqrt{nm/b_{_1}b_{_2}} \cdot c \cdot \alpha(s-m+1) \longrightarrow 0. \tag{C.6}$$

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

$$r = r_{_n} \coloneqq \left\lfloor \frac{\sqrt{nb_{_1}b_{_2}/m}}{c} \right\rfloor, \qquad \ell = \ell_{_n} \coloneqq \left\lfloor \frac{n}{r+s} \right\rfloor, \qquad \vartheta = \vartheta_{_n} \coloneqq s-m+1, \tag{C.7}$$

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

$$\frac{s}{r} \longrightarrow 0; \qquad \ell\alpha(\vartheta) \longrightarrow 0; \qquad \frac{mr}{n} \longrightarrow 0; \qquad \frac{mr}{\ell b_{_1}b_{_2}} \longrightarrow 0; \qquad \frac{m\ell s}{n} \longrightarrow 0. \tag{C.8}$$

Proof. Item (a) will be established by first observing that it is possible to find integers s that ensures that assumption 2.3(f) is compatible with the requirement $m = o\big((nb_{_1}b_{_2})^{\xi}\big)$, for any $\xi \in \big(0, \frac{1}{3}\big)$, and then checking that the exponent $\tau/(2+5\tau) - \lambda$ lies in this interval.

Observe that it is impossible to have $m = o(s)$ and $s = o\Big(\sqrt{nb_{_1}b_{_2}/m}\Big)$ when $m \geq \sqrt{nb_{_1}b_{_2}/m}$, which implies $m < \sqrt{nb_{_1}b_{_2}/m}$, which is equivalent to $m < (nb_{_1}b_{_2})^{1/3}$. Some extra leeway is needed in order to construct the desired integers s, so consider the requirement

$$m = o\big((nb_{_1}b_{_2})^{1/3-\zeta}\big)\,, \qquad \text{for some } \zeta \in \big(0, \tfrac{1}{3}\big)\,. \tag{C.9}$$

Define the integers s by $s \coloneqq m \cdot \mathfrak{s}$, where $\mathfrak{s} \coloneqq 1 \vee \big\lfloor (nb_{_1}b_{_2})^{\zeta/2} \big\rfloor$, and note that this construction ensures that s goes to ∞. Further, $m = o(s)$ holds since $m/s = 1/\mathfrak{s} \to 0$, and $s = o\Big(\sqrt{nb_{_1}b_{_2}/m}\Big)$ holds since

$$\begin{aligned} \frac{s}{\sqrt{nb_{_1}b_{_2}/m}} &\asymp \frac{m \cdot (nb_{_1}b_{_2})^{\zeta/2}}{(nb_{_1}b_{_2}/m)^{1/2}} = \frac{m^{3/2}}{(nb_{_1}b_{_2})^{(1-\zeta)/2}} = \left[\frac{m}{(nb_{_1}b_{_2})^{(1-\zeta)/3}}\right]^{3/2} \\ &= \left[\frac{1}{(nb_{_1}b_{_2})^{2\zeta/3}} \cdot \frac{m}{(nb_{_1}b_{_2})^{1/3-\zeta}}\right]^{3/2} \to \left[\frac{1}{\infty} \cdot 0\right]^{3/2} = 0. \end{aligned} \tag{C.10}$$

This implies that the desired integers s can be found whenever $m = o\big((nb_{_1}b_{_2})^{\xi}\big)$, with $\xi \in \big(0, \frac{1}{3}\big)$. Since the value of $\tau/(2+5\tau) - \lambda$ lies in the interval $\big(0, \frac{1}{5}\big)$, the proof of item (a) is complete.

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

$$s = 1 \vee \left\lfloor \Big(\sqrt{nb_{_1}b_{_2}/m}\Big)^{1-\eta} \right\rfloor, \qquad c = \Big(\sqrt{nb_{_1}b_{_2}/m}\Big)^{\eta/2}, \quad \text{for some } \eta \in (0,1). \tag{C.11}$$

Since $1-\eta$ and $\eta/2$ are in $(0,1)$, it follows from assumption 2.3(b) that s and c goes to ∞ as required. A quick inspection reveals that the product $c \cdot s$ is $o\Big(\sqrt{nb_{_1}b_{_2}/m}\Big)$, proving the first part of eq. (C.6). For the second part of eq. (C.6), keep in mind the similarity with assumption 2.3(d), and observe that c in the limit is asymptotically equivalent to $s^{\eta/2(1-\eta)}$. Since

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 $x \to \infty$, since $x \asymp \lfloor x \rfloor$, that is $\lim x/\lfloor x \rfloor = 1$. Moreover, observe that assumption 2.3(b) implies that n/m goes to ∞. With these observations, all except the last limit in eq. (C.8) are trivial to prove, i.e.

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

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

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

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

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

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

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

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

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

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

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

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

An inspection of the outermost exponent reveals

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

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

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

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

Lemma C.5. *For $K(\boldsymbol{w})$ from definition B.9 (page 51), and $\nu > 2$ from assumption 2.1(b) (page 12), the following holds:*

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

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

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

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

$$
\left|\int_{\mathbb{R}^2} K(w_1, w_2)\, w_1^k w_2^\ell \,\mathrm{d}w_1 \mathrm{d}w_2\right| \leq \int_{\mathbb{R}^2} K(w_1, w_2)\, |w_1^k w_2^\ell| \,\mathrm{d}w_1 \mathrm{d}w_2, \tag{C.16}
$$

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

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

$$
\left|\int_{\mathbb{R}^2} K(w_1, w_2)^2 w_1^k w_2^\ell \,\mathrm{d}w_1 \mathrm{d}w_2\right| \leq \mathcal{C} \left|\int_{\mathbb{R}^2} K(w_1, w_2)\, w_1^k w_2^\ell \,\mathrm{d}w_1 \mathrm{d}w_2\right|, \tag{C.17}
$$

which due to item (a) is finite, thus item (b) holds true.

Next, note that $|K(w_1,w_2)\,w_1^k w_2^\ell|^\nu = |K(w_1,w_2)|^{(\nu-1)}\,|K(w_1,w_2)|\,|w_1^k w_2^\ell|^\nu \le \mathcal{C}^{(\nu-1)}K(w_1,w_2)\,|w_1^k w_2^\ell|^\nu$, which gives the following inequality,

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

from which it is clear that a proof of the finiteness of the right hand side of eq. (C.18) will imply item (c). Since the region of integration can be divided into $\mathcal{A}_{k\ell} = \{\boldsymbol{w} : |w_1^k w_2^\ell| \le 1\}$ and $\mathcal{A}_{k\ell}^c = \mathbb{R}^2 \setminus \mathcal{A}_{k\ell}$, it follows from the non-negativeness of $K(\boldsymbol{w})$, and eqs. (B.27a) and (B.27d) of definition B.9, that

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

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

where the last inequality follows since the assumption $k+\ell \le 2$ ensures that $k\lceil\nu\rceil + \ell\lceil\nu\rceil \le 2\lceil\nu\rceil$. The expression in eq. (C.18) is thus finite – and, as stated in item (c), $K(w_1,w_2)\,w_1^k w_2^\ell \in L^\nu$. □

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

(a) $\int_{\mathbb{R}^2} \sqrt{b_1 b_2} K_{h:\boldsymbol{b}}(\boldsymbol{\zeta}-\boldsymbol{v})\,u_{hq:\boldsymbol{b}}(\boldsymbol{\zeta})\,\mathrm{d}\boldsymbol{\zeta} = O\big(\sqrt{b_1 b_2}\big)$.

(b) $\left(\int_{\mathbb{R}^2} \left|\sqrt{b_1 b_2} K_{h:\boldsymbol{b}}(\boldsymbol{\zeta}-\boldsymbol{v})\,u_{hq:\boldsymbol{b}}(\boldsymbol{\zeta})\right|^\nu \mathrm{d}\boldsymbol{\zeta}\right)^{1/\nu} = O\big(|b_1 b_2|^{(2-\nu)/2\nu}\big)$.

(c) Let $\mathcal{K}_{qr,hj:\boldsymbol{b}}(\boldsymbol{\zeta}_1,\boldsymbol{\zeta}_2) \coloneqq K_{h:\boldsymbol{b}}(\boldsymbol{\zeta}_1-\boldsymbol{v})\,K_{j:\boldsymbol{b}}(\boldsymbol{\zeta}_2-\boldsymbol{v})\,u_{hq:\boldsymbol{b}}(\boldsymbol{\zeta}_1)\,u_{jr:\boldsymbol{b}}(\boldsymbol{\zeta}_2)$, where $\boldsymbol{\zeta}_1$ and $\boldsymbol{\zeta}_2$ either coincide completely (bivariate), have one common component (trivariate), or have no common components (tetravariate). Let κ be the number of variates, and let $\mathrm{d}\boldsymbol{\zeta}(\kappa)$ represent the corresponding κ-variate differential. Then,

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

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

$$\int_{\mathbb{R}^2} \sqrt{b_1 b_2}\cdot\frac{1}{b_1 b_2} K\left(\frac{\zeta_1 - v_1}{b_1}, \frac{\zeta_2 - v_2}{b_2}\right) u_{hq:\boldsymbol{b}}(\zeta_1,\zeta_2)\,\mathrm{d}\zeta_1\,\mathrm{d}\zeta_2, \tag{C.20}$$

which implies that the substitutions $w_1 = (\zeta_1 - v_1)/b_1$ and $w_2 = (\zeta_1 - v_2)/b_2$ gives the integral

$$\begin{aligned}&\int_{\mathbb{R}^2} \frac{\sqrt{b_1 b_2}}{b_1 b_2} K(w_1, w_2)\, u_{hq:\boldsymbol{b}}(b_1 w_1 + v_1, b_2 w_2 + v_2)\, (b_1 \mathrm{d}w_1)\, (b_2 \mathrm{d}w_2) \\ &= \sqrt{b_1 b_2} \cdot \int_{\mathbb{R}^2} K(w_1, w_2)\, u_{hq:\boldsymbol{b}}(b_1 w_1 + v_1, b_2 w_2 + v_2)\, \mathrm{d}w_1 \mathrm{d}w_2. \end{aligned} \tag{C.21}$$

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

$$u_{hq:\boldsymbol{b}}(v_1, v_2) + b_1 c_1 w_1 + b_2 c_2 w_2 + b_1^2 c_{11} w_1^2 + b_1 b_2 c_{12} w_1 w_2 + b_2^2 c_{22} w_2^2, \tag{C.22}$$

for suitable constants c_1, c_2, c_{11}, c_{12} and c_{22}. The integral in eq. (C.21) can thus be expressed as a sum of integrals like those occurring in lemma C.5(a), all of which are finite. The dominant term becomes $O\big(\sqrt{b_1 b_2}\big)$ when $\boldsymbol{b} \to \mathbf{0}^+$, and the conclusion of item (a) follows.

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

$$\begin{aligned}&\left(\int_{\mathbb{R}^2} \left| \sqrt{b_1 b_2} \cdot \frac{1}{b_1 b_2} K(w_1, w_2)\, u_{hq:\boldsymbol{b}}(b_1 w_1 + v_1, b_2 w_2 + v_2) \right|^{\nu} (b_1 \mathrm{d}w_1)\, (b_2 \mathrm{d}w_2) \right)^{1/\nu} \\ &= |b_1 b_2|^{(2-\nu)/2\nu} \left(\int_{\mathbb{R}^2} \left| K(w_1, w_2)\, u_{hq:\boldsymbol{b}}(b_1 w_1 + v_1, b_2 w_2 + v_2) \right|^{\nu} \mathrm{d}w_1 \mathrm{d}w_2 \right)^{1/\nu}. \end{aligned} \tag{C.23}$$

Note that this represent the norm in L^{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

$$|b_1 b_2|^{(2-\nu)/2\nu} \left(u_{hq:\boldsymbol{b}}(v_1, v_2) + b_1 e_1 w_1 + b_2 e_2 w_2 + b_1^2 e_{11} w_1^2 + b_1 b_2 e_{12} w_1 w_2 + b_2^2 e_{22} w_2^2 \right), \tag{C.24}$$

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

The investigation of item (c) requires different substitutions depending on the κ for the configuration under investigation. Noting that the integrand in addition to the scaling factor $b_1 b_2$ always contains the product $K_{h:\boldsymbol{b}}(\boldsymbol{\zeta}_1 - \boldsymbol{v})\, K_{j:\boldsymbol{b}}(\boldsymbol{\zeta}_2 - \boldsymbol{v})$, it follows that it regardless of the value of 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 $b_1 b_2\, \mathrm{d}w_1 \mathrm{d}w_2$ when $\kappa = 2$, $b_1^2 b_2\, \mathrm{d}w_1 \mathrm{d}w_2 \mathrm{d}w_3$ or $b_1 b_2^2\, \mathrm{d}w_1 \mathrm{d}w_2 \mathrm{d}w_3$ when $\kappa = 3$, and $b_1^2 b_2^2\, \mathrm{d}w_1 \mathrm{d}w_2 \mathrm{d}w_3 \mathrm{d}w_4$ when $\kappa = 4$.

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

$$\int_{\mathbb{R}^2} K(w_1, w_2)^2 \cdot \mathcal{U}(w_1, w_2)\, \mathrm{d}w_1 \mathrm{d}w_2, \tag{C.25}$$

where $\mathcal{U}(w_1, w_2)$ is a product whose factors both are of the form encountered in eq. (C.22), i.e. it will be a quartic polynomial in the variables $(b_1 w_1)$ and $(b_2 w_2)$, and its constant term will

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

For the trivariate case, the overlap between $\boldsymbol{\zeta}_1$ and $\boldsymbol{\zeta}_2$ will belong to one of the following configurations, (i) $\boldsymbol{\zeta}_1 = (\zeta_1, \zeta_2)$ and $\boldsymbol{\zeta}_2 = (\zeta_1, \zeta_3)$, (ii) $\boldsymbol{\zeta}_1 = (\zeta_1, \zeta_2)$ and $\boldsymbol{\zeta}_2 = (\zeta_3, \zeta_1)$, (iii) $\boldsymbol{\zeta}_1 = (\zeta_1, \zeta_2)$ and $\boldsymbol{\zeta}_2 = (\zeta_2, \zeta_3)$, or (iv) $\boldsymbol{\zeta}_1 = (\zeta_1, \zeta_2)$ and $\boldsymbol{\zeta}_2 = (\zeta_3, \zeta_2)$. The reasoning is identical for the four cases, so it is sufficient to consider case (i), which gives the following product of kernel functions in the original integral,

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

When the substitution

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

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

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

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 $\ell \in \{0, 1, 2\}$,

$$\begin{aligned}
\int_{\mathbb{R}^1} \mathcal{K}\,(w_1, w_2, w_3)\, w_3^\ell \,\mathrm{d}w_3 &= K(w_1, w_2) \cdot \int_{\mathbb{R}^1} K([(b_2 w_2 + v_2) - v_1]\,/b_1, w_3)\, w_3^\ell \,\mathrm{d}w_3 && \text{(C.29a)}\\
&= K(w_1, w_2) \cdot \mathcal{K}_{2:\ell}([(b_2 w_2 + v_2) - v_1]\,/b_1) && \text{(C.29b)}\\
&\leq \mathcal{D}_{2:\ell} \cdot K(w_1, w_2)\,, && \text{(C.29c)}
\end{aligned}$$

where $\mathcal{D}_{2:\ell}$ is a constant that bounds the function $\mathcal{K}_{2:\ell}$.

Since the substitution in eq. (C.27) transforms the integral of interest into

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

where $\mathcal{U}(w_1, w_2, w_3)$ is a quadratic polynomial in the variables $(b_1 w_1)$ and $(b_2 w_3)$, and a quartic polynomial in w_2 (with coefficients having suitable powers of b_1 and b_2 as factors), the observation in eq. (C.29) implies that an iterated approach to the integral (starting with the w_3-variable) can be used to show that each part of the sum will be bounded by a constant times an integral of the form encountered in lemma C.6(a). The trivariate integral in item (c) can thus be bounded by a sum of finite integrals having coefficients based on powers of b_1 and b_2. From the b_2 factor in eq. (C.30), it follows that the trivariate integral in this case is $O(b_2)$ when $\boldsymbol{b} \to \mathbf{0}^+$. Note that $w_2 = (\zeta_2 - v_1)/b_1$ could have been used as an alternative substitution in eq. (C.27), which by the obvious modifications of the arguments implies that the integral also will be $O(b_1)$ when $\boldsymbol{b} \to \mathbf{0}^+$ – and from this if follows that the integral is $O(b_1 \wedge b_2)$, which completes the proof for the $\kappa = 3$ case of item (c).

The case $\kappa = 4$ is quite simple, since no common components in $\boldsymbol{\zeta}_1$ and $\boldsymbol{\zeta}_2$ implies that the tetravariate integral, after the obvious substitution, corresponds to an expression of the form

$$b_1 b_2 \left(\int_{\mathbb{R}^2} K(\boldsymbol{w}) u_{hq:\boldsymbol{b}}(\boldsymbol{\zeta}(\boldsymbol{w}))\, \mathrm{d}\boldsymbol{w} \right) \cdot \left(\int_{\mathbb{R}^2} K(\boldsymbol{w}) u_{jr:\boldsymbol{b}}(\boldsymbol{\zeta}(\boldsymbol{w}))\, \mathrm{d}\boldsymbol{w} \right), \tag{C.31}$$

where $\boldsymbol{\zeta}(\boldsymbol{w}) = (b_1 w_1 + v_1, b_2 w_2 + v_2)$. The integrals occurring in this product are similar to those encountered in the bivariate case discussed above, and it is clear that the result will be $O(b_1 b_2)$ when $\boldsymbol{b} \to \mathbf{0}^+$, which concludes the proof of item (c). □

Note that the bivariate case of lemma C.6(c) only considers the configuration where the components of $\boldsymbol{\zeta}_1$ and $\boldsymbol{\zeta}_2$ coincide completely, while the configuration where $\boldsymbol{\zeta}_1 = (\zeta_1, \zeta_2)$ and $\boldsymbol{\zeta}_2$ is the diagonal reflection (ζ_2, ζ_1) has been left out. This restriction does not pose a problem for the asymptotic investigation of $\widehat{f}^{m}_{\boldsymbol{v}|p}(\omega)$ when the point $\boldsymbol{v} = (v_1, v_2)$ lies upon the diagonal, i.e. when $v_1 = v_2$, since the diagonal folding property ensures that it is sufficient to consider positive lags for the point $\boldsymbol{v}$ in this case. For the general case, where $v_1 \neq v_2$, the following adjusted version of lemma C.6(c) is needed, where one of the kernels use $\boldsymbol{v}$ and the other use the diagonally reflected point $\breve{\boldsymbol{v}} = (v_2, v_1)$.

Lemma C.7. *The following holds for $u_{hq:\boldsymbol{b}}(\boldsymbol{w})$ and $K_{h:\boldsymbol{b}}(\boldsymbol{y}_h - \boldsymbol{v})$ from definitions B.7 and B.9, when the point $\boldsymbol{v} = (v_1, v_2)$ does not coincide with its diagonal reflection $\breve{\boldsymbol{v}} = (v_2, v_1)$, i.e. $v_1 \neq v_2$.*
Let $\mathcal{K}_{qr,hj:\boldsymbol{b}}(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2; \boldsymbol{v}, \breve{\boldsymbol{v}}) \coloneqq K_{h:\boldsymbol{b}}(\boldsymbol{\zeta}_1 - \boldsymbol{v})\, K_{j:\boldsymbol{b}}(\boldsymbol{\zeta}_2 - \breve{\boldsymbol{v}})\, u_{hq:\boldsymbol{b}}(\boldsymbol{\zeta}_1)\, u_{jr:\boldsymbol{b}}(\boldsymbol{\zeta}_2)$, where $\boldsymbol{\zeta}_1$ and $\boldsymbol{\zeta}_2$ either are diagonal reflections of each other (bivariate), have one common component (trivariate), or have no common components (tetravariate). Let κ be the number of variates, and let $\mathrm{d}\boldsymbol{\zeta}(\kappa)$ represent the corresponding κ-variate differential. Then,

$$\int_{\mathbb{R}^\kappa} (b_1 b_2)\, \mathcal{K}_{qr,hj:\boldsymbol{b}}(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2; \boldsymbol{v}, \breve{\boldsymbol{v}})\, \mathrm{d}\boldsymbol{\zeta}(\kappa) = \begin{cases} o(1) & \kappa = 2, \\ O(b_1 \wedge b_2) & \kappa = 3, \\ O(b_1 b_2) & \kappa = 4. \end{cases}$$

Proof. The statements for the trivariate and tetravariate cases are identical to those in lemma C.6(c), and so are the proofs, i.e. the same substitutions can be applied for the present cases of interest.

For the bivariate case, the substitution $w_1 = (\zeta_1 - v_1)/b_1$ and $w_2 = (\zeta_1 - v_2)/b_2$ gives that the integral $\int_{\mathbb{R}^2} K(w_1, w_2)^2 \cdot \mathcal{U}(w_1, w_2)\, \mathrm{d}w_1\, \mathrm{d}w_2$ from eq. (C.25) is replaced with a sum of integrals of the form,

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

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

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 $K(\boldsymbol{w})$ is the product normal kernel. The observation that the bivariate case of lemma C.7 is $o(1)$ can also be derived from the realisation that $K_{h:\boldsymbol{b}}(\boldsymbol{\zeta}_1 - \boldsymbol{v})$ and $K_{j:\boldsymbol{b}}(\boldsymbol{\zeta}_2 - \breve{\boldsymbol{v}})$ are entities that converge towards two different bivariate Dirac delta functions, and the limit of the integral becomes zero since these delta functions sifts out different points.

C.5 A few details related to 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

$$\mathcal{F}_t^s \coloneqq \sigma(Y_t, \ldots, Y_s)\,, \tag{C.33}$$

where t and s are allowed to take the values $-\infty$ and $+\infty$ respectively.

Note in particular, that if a new random variable is defined by means of a measurable function $\xi(\boldsymbol{y}_{\bar{m}})$ from $\mathbb{R}^{m+1}$ to $\mathbb{R}$, i.e. $\mathcal{Y}_{m:t} \coloneqq \xi(\boldsymbol{Y}_{\bar{m}:t})$, then $\mathcal{Y}_{m:t} \in \mathcal{F}_t^{t+m}$.

Inheritance of α-mixing

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

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

from which it is an easy task to verify that a derived process, like the $\mathcal{Y}_{m:t}$ mentioned above, will have an inherited 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

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

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]

[34] These 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 $L^{\nu} \coloneqq L^{\nu}(\Omega, \mathcal{G}, \mu)$ is defined to be the class of measurable real functions ζ for which $|\zeta|^{\nu}$ is integrable, that is,

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

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

$$L_h^{\nu} \quad \text{— the } L^{\nu} \text{ spaces related to the densities } g_h, \tag{C.38a}$$
$$L_{\overline{m}}^{\nu} \quad \text{— the } L^{\nu} \text{ space related to the density } g_{\overline{m}}. \tag{C.38b}$$

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

$$\|\zeta(z)\|_{\nu} \coloneqq \left(\int_{\Omega} |\zeta(z)|^{\nu} \, \mathrm{d}\mu \right)^{1/\nu} = \left(\mathrm{E}[|\zeta(Z)|^{\nu}] \right)^{1/\nu} \tag{C.39}$$

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

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

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

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

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

$$\begin{gathered} \textit{If } \mathrm{E}[|\zeta_h(\boldsymbol{Y}_{h:t})|^{\nu}] < \infty \textit{ for } h = 1, \ldots, m, \textit{ then} \\ \left(\mathrm{E}\left[\left| \textstyle\sum_{h=1}^{m} a_h \zeta_h(\boldsymbol{Y}_{h:t}) \right|^{\nu} \right] \right)^{1/\nu} \leq \textstyle\sum_{h=1}^{m} |a_h| \left(\mathrm{E}[|\zeta_h(\boldsymbol{Y}_{h:t})|^{\nu}] \right)^{1/\nu} < \infty. \end{gathered}$$

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

C.6 The one-parameter local Gaussian approximation

The input parameter p in the local Gaussian spectral density $f_{\boldsymbol{v}|p}(\omega)$ 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 $\psi_5(\boldsymbol{w})$ or $\psi_1(\boldsymbol{w})$, where

$$\psi_5(\boldsymbol{w};\mu_1,\mu_2,\sigma_1,\sigma_2,\rho) \coloneqq \frac{1}{2\pi\cdot\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{\sigma_1^2\big(w_1-\mu_1\big)^2-2\sigma_1\sigma_2\rho\big(w_1-\mu_1\big)\big(w_2-\mu_2\big)+\sigma_2^2\big(w_2-\mu_2\big)^2}{2\sigma_1^2\sigma_2^2(1-\rho^2)}\right\}, \tag{C.42a}$$

$$\psi_1(\boldsymbol{w};\rho) \coloneqq \frac{1}{2\pi\cdot\sqrt{1-\rho^2}}\exp\left\{-\frac{w_1^2-2\rho w_1 w_2+w_2^2}{2(1-\rho^2)}\right\}. \tag{C.42b}$$

As noted in section 2.1.3, the simplified approach based on $\psi_1(\boldsymbol{w})$ 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 $\rho_{\boldsymbol{v}|1}$ gave more or less inconclusive results whereas those based on $\rho_{\boldsymbol{v}|5}$ 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 $\psi_1(\boldsymbol{w})$, i.e. how appropriate will it in general be to use the estimated local correlations obtained from $\psi_1(\boldsymbol{w})$. It is e.g. evident from eq. (C.42) that $\psi_1(\boldsymbol{w})$ lacks the flexibility of $\psi_5(\boldsymbol{w})$, 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 $\psi_1(\boldsymbol{w})$ 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 $\boldsymbol{v}$, the target of interest moves from the estimated density $\widehat{\psi}_1(\boldsymbol{v})$ to the estimated parameter $\widehat{\rho}_{\boldsymbol{v}|1}$.

The analysis related to $\psi_1(\boldsymbol{w})$ is straightforward to do, since it in essence only requires that some integrals containing the factor $K_{\boldsymbol{b}}(\boldsymbol{w}-\boldsymbol{v})$ 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 $\psi_1(\boldsymbol{w})$ 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

$$\phi_{2}(w;\mu,\sigma) \coloneqq \tfrac{1}{\sqrt{2\pi}\cdot\sigma}\exp\left\{-\tfrac{(w-\mu)^2}{2\sigma^2}\right\}, \tag{C.43a}$$

$$\phi_{1}(w;\sigma) \coloneqq \tfrac{1}{\sqrt{2\pi}\cdot\sigma}\exp\left\{-\tfrac{w^2}{2\sigma^2}\right\}, \tag{C.43b}$$

where $\phi_{1}(w;\sigma)$, together with its score function $u_{1}(w;\sigma) = \frac{d}{\mathrm{d}\sigma}\log\phi_{1}(w;\sigma) = (w^2-\sigma^2)/\sigma^3$, is the target of interest for the present investigation. It is clear from the rigidity of the one-parameter alternative $\phi_{1}(\boldsymbol{w};\sigma)$ 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 $b \to 0^{+}$ limit of the penalty functions

$$q_{v|b|i} = \int K_{b}(w-v)[\phi_{i}(w;\sigma) - g(w)\cdot\log\phi_{i}(w;\sigma)]\,\mathrm{d}w, \tag{C.44}$$

where $K_{b}(w-v) \coloneqq \frac{1}{b}\cdot K(\frac{w-v}{b})$, 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 $\frac{\mathrm{d}}{\mathrm{d}\sigma}q_{v|b|1} = 0$, i.e. the following equation should be satisfied,

$$\int K_{b}(w-v)u_{1}(w;\sigma_{v|b|1})\big[\phi_{1}(w;\sigma_{v|b|1}) - g(w)\big]\,\mathrm{d}w = 0. \tag{C.45}$$

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

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

where R is the remainder term of the Taylor expansion.

It follows from eq. (C.46) that the $b \to 0^{+}$ 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

$$u_{1}(v;\sigma_{v|b|1}) = \frac{b^2\cdot\left\{\frac{1}{2}h^{(2)}(v;\sigma_{v|b|1}) + \int R(v+bz;b,\sigma_{v|b|1})z^2K(z)\,\mathrm{d}z\right\}}{\big(g(v) - \phi_{1}(v;\sigma_{v|b|1})\big)}, \tag{C.47}$$

and that the $b \to 0^{+}$ 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 $n \to \infty$ limit of the matrices from the Klimko-Nelson approach (see theorem B.1) are given as the following scalar expressions,

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

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 $b \to 0^+$ and $n \to \infty$ limit of the asymptotic variance of $\sqrt{n}\left(\widehat{\sigma}_{v|b|1} - \sigma_{v|1}\right)$ is thus based on the $b \to 0^+$ limit of

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

from which it follows that the asymptotic variance of $\sqrt{nb}\left(\widehat{\sigma}_{v|b|1} - \sigma_{v|1}\right)$ 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 $b \to 0^+$ 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 $\widehat{\sigma}_{v|b|1}$ converge slower in these cases, but $\widehat{\sigma}_{v|b|1}$ 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 $\psi_1(\boldsymbol{w}; \rho)$

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 $\psi_1(\boldsymbol{w}; \rho)$ can attain any positive value when $\boldsymbol{w} = (w_1, w_2)$ 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

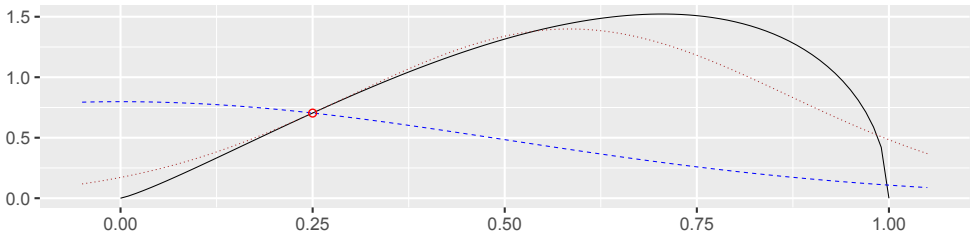

Figure 15: Simplified univariate investigation based on 'local Gaussian standard deviations'.

points $\boldsymbol{w}$ than those already mentioned, an inspection of the corresponding score function,

$$u_1(\boldsymbol{w};\rho) = \frac{\mathrm{d}}{\mathrm{d}\rho}\log\left(\psi_1(\boldsymbol{w};\rho)\right) = \frac{-\rho^3 + \rho^2 w_1 w_2 + \rho\left(1-(w_1^2+w_2^2)\right) + w_1 w_2}{\left(1-\rho^2\right)^2}, \tag{C.50}$$

together with the requirement that ρ lies in $(-1,1)$, gives that there is an upper limit $\psi_1(\boldsymbol{w};\rho^\circ_{\boldsymbol{w}|1})$ for $\psi_1(\boldsymbol{w};\rho)$ at this particular point. Note that $\rho^\circ_{\boldsymbol{w}|1}$ is a solution of $u_1(\boldsymbol{w};\rho)=0$, 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 $\psi_1(\boldsymbol{w})$. The observations below follows from those given in remarks C.5 and C.6. Not that the bivariate target density $g(\boldsymbol{w})$ here corresponds to a copula-model with standard normal margins, and $\boldsymbol{v}=(v_1,v_2)$ represents the point at which the density estimation and extraction of correlation $\rho_{\boldsymbol{v}|1}$ is performed.

(a) If $\boldsymbol{v}=\boldsymbol{0}$ and $g(\boldsymbol{0}) < 1/2\pi$, then $\rho_{\boldsymbol{v}|1}=0$ and the estimated density always becomes $\psi_1(\boldsymbol{0};0)=1/2\pi$. The estimates of $\rho_{\boldsymbol{v}|1}$ will in this case converge towards $\rho_{\boldsymbol{v}|1}=0$ with a slower convergence rate than $\sqrt{nb_1b_2}$, and $\rho_{\boldsymbol{v}|1}=0$ will always occur when $g(\boldsymbol{0})<1/2\pi$, regardless of how $g(\boldsymbol{w})$ behaves in a neighbourhood of $\boldsymbol{v}=\boldsymbol{0}$.

(b) If $\boldsymbol{v}$ 'lies off the diagonals', i.e. $|v_1|\neq|v_2|$, and $g(\boldsymbol{v}) > \psi_1(\boldsymbol{v};\rho^\circ_{\boldsymbol{v}|1})$, then $\rho_{\boldsymbol{v}|1}=\rho^\circ_{\boldsymbol{v}|1}$ and the estimated density always becomes $\psi_1(\boldsymbol{v};\rho^\circ_{\boldsymbol{v}|1})$. The estimates of $\rho_{\boldsymbol{v}|1}$ will in this case converge towards $\rho^\circ_{\boldsymbol{v}|1}$ with a slower convergence rate than $\sqrt{nb_1b_2}$, and $\rho_{\boldsymbol{v}|1}=\rho^\circ_{\boldsymbol{v}|1}$ will always occur when $g(\boldsymbol{v}) > \psi_1(\boldsymbol{v};\rho^\circ_{\boldsymbol{v}|1})$, regardless of how $g(\boldsymbol{w})$ behaves in a neighbourhood of $\boldsymbol{v}$.

(c) The one-parameter approximation $\psi_1(\boldsymbol{w};\rho)$ will give the correct density estimate of $g(\boldsymbol{w})$ at $\boldsymbol{v}$ for other cases than those excluded above, but the corresponding correlation $\rho_{\boldsymbol{v}|1}$ might differ substantially from the one obtained from the five-parameter approximation.

The case in item (b) can e.g. be observed when $g(\boldsymbol{w})$ is the probability density function of a Clayton copula with standard normal marginals.

Remark C.7. The preceding discussion assumed that the target density $g(\boldsymbol{w})$ 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 $f_{\boldsymbol{v}|5}(\omega)$ and $f_{\boldsymbol{v}|1}(\omega)$ 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 $\boldsymbol{b} = (0.5, 0.5)$. As can be seen, the estimates based on the $\psi_1(\boldsymbol{w})$-approximation tend to have the same sign as those from the $\psi_5(\boldsymbol{w})$-approximation, but the values are in general much closer to 0.

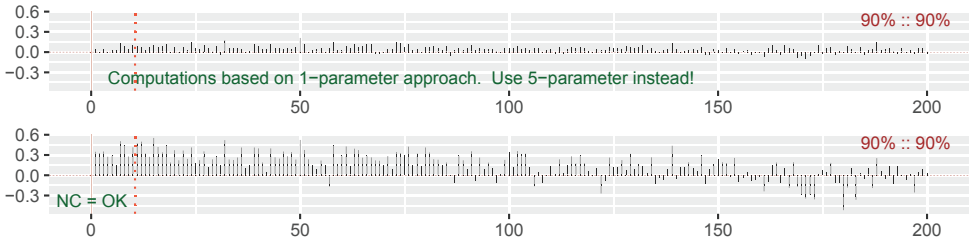

Figure 16: A comparison of estimated local Gaussian autocorrelations from the `dmbp`-data: $\widehat{\rho}_{\boldsymbol{v}|1}(h)$ from the $\psi_1(\boldsymbol{w})$-approximation (top) versus $\widehat{\rho}_{\boldsymbol{v}|5}(h)$ from the $\psi_5(\boldsymbol{w})$-approximation (bottom).

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 $g_h(\boldsymbol{y}_h)$ of $\left(Y_{t+h}, Y_t\right)$ 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 $g_h(\boldsymbol{y}_h)$ behaves in such a manner that the one-parameter approximation $\psi_1(\boldsymbol{w}; \rho)$ 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 $\left(Y_{t+h}, Y_t\right)$, and 2) the selected copula-models (with standard normal marginals) are compared against the one-parameter case $\psi_1(\boldsymbol{w}; \rho)$.

Remark C.10. Figures 14 and 16 are the only plots in this paper that compares the results based on correlations extracted from $\psi_1(\boldsymbol{w})$ with those based on correlations from $\psi_5(\boldsymbol{w})$. 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, Aniversary 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/092753989390006D

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-validatory model-selection for dependent data: hv-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

Paper II

2.2 Nonlinear cross-spectrum analysis via the local Gaussian correlation

Lars Arne Jordanger and Dag Tjøstheim

Nonlinear cross-spectrum analysis via the local Gaussian correlation

Lars Arne Jordanger Dag Tjøstheim

Abstract

Spectrum analysis can detect frequency related structures in a time series $\{\boldsymbol{Y}_t\}_{t\in\mathbb{Z}}$, but may in general be an inadequate tool if asymmetries or other nonlinear phenomena are present. This limitation is a consequence of the way the spectrum is based on the second order moments (auto and cross-covariances), and alternative approaches to spectrum analysis have thus been investigated based on other measures of dependence. One such approach was developed for univariate time series in Jordanger and Tjøstheim (2017), where it was seen that a *local Gaussian autospectrum* $f_{\boldsymbol{v}}(\omega)$, based on the *local Gaussian autocorrelations* $\rho_{\boldsymbol{v}}(h)$ from Tjøstheim and Hufthammer (2013), could detect local structures in time series that looked like white noise when investigated by the ordinary auto-spectrum $f(\omega)$. The *local Gaussian approach* in this paper is extended to a *local Gaussian cross-spectrum* $f_{k\ell:\boldsymbol{v}}(\omega)$ for multivariate time series. The local cross-spectrum $f_{k\ell:\boldsymbol{v}}(\omega)$ has the desirable property that it coincides with the ordinary cross-spectrum $f_{k\ell}(\omega)$ for Gaussian time series, which implies that $f_{k\ell:\boldsymbol{v}}(\omega)$ can be used to detect non-Gaussian traits in the time series under investigation. In particular: If the ordinary spectrum is flat, then peaks and troughs of *the local Gaussian spectrum* can indicate nonlinear traits, which potentially might discover *local periodic phenomena* that goes undetected in an ordinary spectral analysis.

1 Introduction

The auto and cross-covariances $\big\{\{\gamma_{k\ell}(h)\}_{h\in\mathbb{Z}}\big\}_{k,\ell=1}^{d}$ from a time series $\big\{\boldsymbol{Y}_t=\big(Y_{1,t},\ldots,Y_{d,t}\big)\big\}_{t\in\mathbb{Z}}$, can range from determining it completely (Gaussian time series) to containing no information at all (GARCH-type models). The auto- and cross-spectral densities $\{f_{k\ell}(\omega)\}_{k,\ell=1}^{d}$ (based on these second order moments) inherit these features, and they may thus be inadequate tools when the task of interest is to investigate non-Gaussian time series containing asymmetries or other nonlinear structures – like those observed in stock returns, cf. e.g. Hong et al. (2007).

In ordinary spectral analysis, if $\big\{Y_{k,t}\big\}_{t\in\mathbb{Z}}$ and $\big\{Y_{\ell,t}\big\}_{t\in\mathbb{Z}}$ are jointly weakly stationary, and if the cross-covariances $\gamma_{k\ell}(h) := \mathrm{Cov}\big(Y_{k,t+h},Y_{\ell,t}\big)$ are absolutely summable, then the cross-spectrum $f_{k\ell}(\omega)$ is defined as the Fourier transform of the autocovariances, .i.e.

$$f_{k\ell}(\omega)=\sum_{h\in\mathbb{Z}}\gamma_{k\ell}(h)\cdot e^{-2\pi i\omega h}. \tag{1.1}$$

The expression for the inverse Fourier transform reveals, when $h = 0$, that the covariance $\text{Cov}\big(Y_{k,t}, Y_{\ell,t}\big) = \gamma_{k\ell}(0)$ can be expressed as the integral $\int_{-1/2}^{1/2} f_{k\ell}(\omega)\,\mathrm{d}\omega$. This makes it possible to inspect how the interaction between the marginal time series varies with the frequency ω. An inspection of the cross-spectrum $f_{k\ell}(\omega)$ is a bit more complicated than that of the auto-spectrum, since $f_{k\ell}(\omega)$ in general will be a complex-valued function. It is thus usually the following real valued functions that are investigated,

$$c_{k\ell}(\omega) = \text{Re}\left(f_{k\ell}(\omega)\right), \qquad q_{k\ell}(\omega) = -\,\text{Im}\left(f_{k\ell}(\omega)\right), \tag{1.2a}$$

$$\alpha_{k\ell}(\omega) = \text{Mod}\left(f_{k\ell}(\omega)\right), \qquad \phi_{k\ell}(\omega) = \quad \text{Arg}\left(f_{k\ell}(\omega)\right), \tag{1.2b}$$

where $c_{k\ell}(\omega)$, $q_{k\ell}(\omega)$, $\alpha_{k\ell}(\omega)$ and $\phi_{k\ell}(\omega)$, respectively, are referred to as the cospectrum, quadrature spectrum, amplitude spectrum and phase spectrum. Note that $c_{k\ell}(\omega)$ always integrates to one over one period, whereas $q_{k\ell}(\omega)$ always integrates to zero.

The *coherence* $\mathcal{K}_{k\ell}(\omega) \coloneqq f_{k\ell}(\omega)/\sqrt{f_{kk}(\omega)f_{\ell\ell}(\omega)}$ is an important tool when a spectral analysis is performed on a multivariate time series, in particular since $\mathcal{K}_{k\ell}(\omega)$ can be realised as the correlation of $\mathrm{d}Z_k(\omega)$ and $\mathrm{d}Z_\ell(\omega)$, where $Z_k(\omega)$ and $Z_\ell(\omega)$ are the right continuous orthogonal-increment processes that by the Spectral Representation Theorem correspond to the weakly stationary time series $\big\{Y_{k,t}\big\}_{t\in\mathbb{Z}}$ and $\big\{Y_{\ell,t}\big\}_{t\in\mathbb{Z}}$, see e.g. Brockwell and Davis (1986, p. 436) for details. The *squared coherence* $|\mathcal{K}_{k\ell}(\omega)|^2$ is of interest since its value (in the interval $[0,1]$) reveals to what extent the two time series $\big\{Y_{k,t}\big\}_{t\in\mathbb{Z}}$ and $\big\{Y_{\ell,t}\big\}_{t\in\mathbb{Z}}$ can be related by a linear filter.

Other spectral approaches, involving different generalisations of the auto-spectrum $f(\omega)$ were discussed in Jordanger and Tjøstheim (2017, section 1), and the majority of the approaches were based on the following idea: The second order moments captured by the autocovariances $\{\gamma(h)\}_{h\in\mathbb{Z}}$ can be replaced by alternative dependence measures ξ_h computed from the bivariate random variables $\big(Y_{t+h}, Y_t\big)$, and a spectral density approach can then (under suitable regularity conditions) be defined as the Fourier transform of $\{\xi_h\}_{h\in\mathbb{Z}}$. For multivariate time series, the natural extension is then to define similar measures $\xi_{k\ell:h}$ for the bivariate random variables $\big(Y_{k,t+h}, Y_{\ell,t}\big)$, and then use the corresponding Fourier-transform as an alternative to the cross-spectrum $f_{k\ell}(\omega)$.

It does not seem to be the case (yet) that multivariate versions have been investigated for all of the possible generalisations of the auto-spectrum $f(\omega)$, but some generalisations do exist. The first extension of the cross-spectrum $f_{k\ell}(\omega)$ along these lines is the *polyspectra* introduced in Brillinger (1965), which is the multivariate version of the higher order moments/cumulants approach to spectral analysis, see Brillinger (1984, 1991); Tukey (1959). Another generalisation of $f_{k\ell}(\omega)$ is given in Chung and Hong (2007), where the generalised function approach introduced in Hong (1999) is used to set up a cross-spectrum that can be used for the testing of directional predictability in foreign exchange markets.

A local Gaussian spectral density $f_{\boldsymbol{v}}(\omega)$ for univariate strictly stationary time series was defined in Jordanger and Tjøstheim (2017), based on the local Gaussian auto-correlations $\rho_{\boldsymbol{v}}(h)$ from Tjøstheim and Hufthammer (2013). A simple adjustment gives the local Gaussian cross-correlations $\rho_{k\ell:\boldsymbol{v}}(h)$ for multivariate strictly stationary time series, from which a local Gaussian analogue $f_{k\ell:\boldsymbol{v}}(\omega)$ of the cross-spectrum $f_{k\ell}(\omega)$ can be constructed using the Fourier transform. The local Gaussian version of the cross-spectrum enables local Gaussian alternatives to be defined of the cospectrum, quadrature spectrum, amplitude spectrum, and phase spectrum, by

simply copying the setup used in the ordinary (global) case. Local Gaussian analogues of the coherence and squared coherence were investigated in the preparation for this paper, but then discarded, see remark 2.5 for further details.

An overview of the paper is as follows: Section 2 defines the *local Gaussian cross-spectrum* $f_{k\ell:\boldsymbol{v}}(\omega)$, which immediately gives the related local Gaussian variants of the cospectrum, quadrature spectrum, amplitude spectrum and phase spectrum from eq. (1.2). The asymptotic theory for the estimators are then presented (some technical details and proofs are postponed to the appendices). The real and simulated examples in section 3 shows that estimates of $f_{k\ell:\boldsymbol{v}}(\omega)$ can be used to detect and investigate nonlinear structures in non-Gaussian white noise, and in particular that $f_{k\ell:\boldsymbol{v}}(\omega)$ can detect local periodic phenomena that goes undetected in an ordinary spectral analysis. Note that the scripts needed for the reproduction of these examples are contained in the R-package `localgaussSpec`,[1] where it in addition is possible to use an interactive solution to see how adjustments of the input parameters (used in the estimation algorithms) influence the estimates of $f_{k\ell:\boldsymbol{v}}(\omega)$. A discussion is given in section 4, and section 5 presents the conclusion.

2 Definitions

This section will present the formal definitions of the local Gaussian versions of the cross-correlation $f_{k\ell}(\omega)$ and its derived entities. The details are almost identical to those encountered when the local Gaussian spectral density was introdued in Jordanger and Tjøstheim (2017), and the present discussion will thus only give short summaries of descriptions and arguments already undertaken in (Jordanger and Tjøstheim, 2017).

2.1 The local Gaussian correlations

At the core of the generalisation of eq. (1.1) lies the local Gaussian correlation $\rho_{\boldsymbol{v}}$ from Tjøstheim and Hufthammer (2013). The theoretical treatment can be based directly on $\rho_{\boldsymbol{v}}$, but the numerical convergence of the estimation-algorithm might then sometimes fail, in particular if the samples contains outliers. As noted in (Jordanger and Tjøstheim, 2017, sections 2.1.2 and 2.1.3), this estimation-problem can be countered by the help of the two revised versions $\rho_{\boldsymbol{v}|5}$ and $\rho_{\boldsymbol{v}|1}$, where a *normalisation of the marginals* (see definition 2.1 below) are performed before the estimation-algorithm is used. The numbers 5 and 1 refer to whether the estimation-algorithm use a local Gaussian correlation that originates from a *five-parameter* or a *one-parameter* local Gaussian approximation, as shown below,

$$\psi_5(\boldsymbol{w};\mu_1,\mu_2,\sigma_1,\sigma_2,\rho) \coloneqq \frac{1}{2\pi\cdot\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{\sigma_1^2\big(w_1-\mu_1\big)^2-2\sigma_1\sigma_2\rho\big(w_1-\mu_1\big)\big(w_2-\mu_2\big)+\sigma_2^2\big(w_2-\mu_2\big)^2}{\sigma_1^2\sigma_2^2(1-\rho^2)}\right\}, \tag{2.1a}$$

$$\psi_1(\boldsymbol{w};\rho) \coloneqq \frac{1}{2\pi\cdot\sqrt{1-\rho^2}} \exp\left\{-\frac{w_1^2-2\rho w_1 w_2+w_2^2}{1-\rho^2}\right\}. \tag{2.1b}$$

Both $\rho_{\boldsymbol{v}|5}$ and $\rho_{\boldsymbol{v}|1}$ can be used as the starting point for the theoretical investigation, and the notation $\rho_{\boldsymbol{v}|p}$ will be used to indicate that both of the alternatives are discussed simultaneously. Although both of the alternatives can be used, only the constructions based on $\rho_{\boldsymbol{v}|5}$ will in

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

general be able to properly detect local properties of the investigated time series. The discussion in (Jordanger and Tjøstheim, 2017, appendix C.6)) gives some reasons for the failure of the one-parameter local Gaussian approach in this context.

Remark 2.1. In order to have a unified notation, the two functions in eq. (2.1) will later on be denoted by $\psi_p\big(\boldsymbol{w};\boldsymbol{\theta}_p\big)$, where additional indices will be added to $\boldsymbol{\theta}_p$ in order to identify the targeted density and the point of investigation.

2.2 The local Gaussian cross-spectrum

The definition of the local Gaussian cross-spectrum density is almost identical to the definition of the local Gaussian spectral density from (Jordanger and Tjøstheim, 2017, section 2.2), which in this paper henceforth will be referred to as the local Gaussian auto-spectrum.

Definition 2.1. *For a strictly stationary multivariate time series $\{\boldsymbol{Y}_t\}_{t\in\mathbb{Z}}$, where $\boldsymbol{Y}_t = \big(Y_{1,t},\ldots,Y_{d,t}\big)$, the local Gaussian cross-spectrum of the marginal time series $\big\{Y_{k,t}\big\}_{t\in\mathbb{Z}}$ and $\big\{Y_{\ell,t}\big\}_{t\in\mathbb{Z}}$ is constructed in the following manner.*

(a) With G_k and G_ℓ the univariate marginal cumulative distribution of respectively $\big\{Y_{k,t}\big\}_{t\in\mathbb{Z}}$ and $\big\{Y_{\ell,t}\big\}_{t\in\mathbb{Z}}$, and Φ the cumulative distribution of the univariate standard normal distribution, define normalised versions $\big\{Z_{k,t}\big\}_{t\in\mathbb{Z}}$ and $\big\{Z_{\ell,t}\big\}_{t\in\mathbb{Z}}$ by

$$\big\{Z_{k,t} \coloneqq \Phi^{-1}\big(G_k\big(Y_{k,t}\big)\big)\big\}_{t\in\mathbb{Z}}, \qquad \big\{Z_{\ell,t} \coloneqq \Phi^{-1}\big(G_\ell\big(Y_{\ell,t}\big)\big)\big\}_{t\in\mathbb{Z}}. \tag{2.2}$$

(b) For a given point $\boldsymbol{v} = (v_1, v_2)$ and for each bivariate lag h pair $\boldsymbol{Z}_{k\ell:h:t} \coloneqq \big(Z_{k:t+h}, Z_{\ell:t}\big)$, a local Gaussian cross-correlation $\rho_{k\ell:\boldsymbol{v}|p}(h)$ can be computed, where the p specifies if the correlations stems from a one or a five parameter local Gaussian approximation of the bivariate density of $\boldsymbol{Z}_{k\ell:h:t}$ at (v_1, v_2).

(c) When $\sum_{h\in\mathbb{Z}}\big|\rho_{k\ell:\boldsymbol{v}|p}(h)\big| < \infty$, the local Gaussian cross-spectrum at the point $\boldsymbol{v}$ is defined as

$$f_{k\ell:\boldsymbol{v}|p}(\omega) \coloneqq \sum_{h=-\infty}^{\infty} \rho_{k\ell:\boldsymbol{v}|p}(h)\cdot e^{-2\pi i\omega h}, \qquad \omega\in\left[-\tfrac{1}{2},\tfrac{1}{2}\right]. \tag{2.3}$$

Remark 2.2. The definition of the local Gaussian auto-spectrum is in essence the same as the one given here for the local Gaussian cross-spectrum, with the minor adjustment that $k = \ell$ in the auto-spectrum case – which requires the added convention that $\rho_{kk:\boldsymbol{v}|p}(0) \equiv 1$ for all points $\boldsymbol{v}$.

The basic properties of the local Gaussian cross-spectrum are quite similar to those encountered for the local Gaussian auto-spectrum in (Jordanger and Tjøstheim, 2017, lemma 2.3).

Lemma 2.2. *The following properties holds for $f_{k\ell:\boldsymbol{v}|p}(\omega)$.*

(a) $f_{k\ell:\boldsymbol{v}|p}(\omega)$ coincides with $f_{k\ell}(\omega)$ for all $\boldsymbol{v}\in\mathbb{R}^2$ when $\{\boldsymbol{Y}_t\}_{t\in\mathbb{Z}}$ is a multivariate Gaussian time series.

(b) The following holds when $\breve{\boldsymbol{v}} \coloneqq (v_2, v_1)$ is the diagonal reflection of $\boldsymbol{v} = (v_1, v_2)$;

$$f_{k\ell:\boldsymbol{v}|p}(\omega) = \overline{f_{\ell k:\breve{\boldsymbol{v}}|p}(\omega)}, \tag{2.4a}$$

$$f_{k\ell:\boldsymbol{v}|p}(\omega) = \rho_{k\ell:\boldsymbol{v}|p}(0) + \sum_{h=1}^{\infty}\rho_{\ell k:\breve{\boldsymbol{v}}|p}(h)\cdot e^{+2\pi i\omega h} + \sum_{h=1}^{\infty}\rho_{k\ell:\boldsymbol{v}|p}(h)\cdot e^{-2\pi i\omega h}. \tag{2.4b}$$

Proof. Item (a) follows since the local Gaussian cross-correlations $\rho_{k\ell:\boldsymbol{v}|p}(h)$ by construction coincides with the ordinary (global) cross-correlations $\rho(h)$ in the Gaussian case. For the proof of item (b), the key observation is that the *diagonal folding property* that was observed for the local Gaussian auto-spectrum, see (Jordanger and Tjøstheim, 2017, lemma C.1), extends directly to the present case, i.e. $\rho_{k\ell|\boldsymbol{v}}(-h) = \rho_{\ell k|\breve{\boldsymbol{v}}}(h)$, where $\breve{\boldsymbol{v}} = (v_2, v_1)$ is the *diagonally reflected point corresponding to* $\boldsymbol{v}$. This implies that $f_{k\ell:\boldsymbol{v}|p}(\omega) = \overline{f_{k\ell:\boldsymbol{v}|p}(-\omega)} = \overline{f_{\ell k:\breve{\boldsymbol{v}}|p}(\omega)}$, and it also follows that eq. (2.3) can be reexpressed as eq. (2.4b). □

2.3 Related local Gaussian entities

From the definition of the local Gaussian cross-spectrum, it is possible to define related spectra in the same manner as those mentioned for the ordinary spectrum in eq. (1.2).

Definition 2.3. *The local Gaussian versions of the cospectrum $c_{k\ell}(\omega)$, the quadrature spectrum $q_{k\ell}(\omega)$, the amplitude spectrum $\alpha_{k\ell}(\omega)$ and the phase spectrum $\phi_{k\ell}(\omega)$, are given by*

$$c_{k\ell:\boldsymbol{v}|p}(\omega) := \operatorname{Re}\left(f_{k\ell:\boldsymbol{v}|p}(\omega)\right) = \rho_{k\ell:\boldsymbol{v}|p}(0) + \sum_{h=1}^{\infty} \cos\left(2\pi\omega h\right)\left[\rho_{k\ell:\boldsymbol{v}|p}(h) + \rho_{k\ell:\breve{\boldsymbol{v}}|p}(h)\right], \tag{2.5a}$$

$$q_{k\ell:\boldsymbol{v}|p}(\omega) := -\operatorname{Im}\left(f_{k\ell:\boldsymbol{v}|p}(\omega)\right) = \sum_{h=1}^{\infty} \sin\left(2\pi\omega h\right)\left[\rho_{k\ell:\boldsymbol{v}|p}(h) - \rho_{k\ell:\breve{\boldsymbol{v}}|p}(h)\right], \tag{2.5b}$$

$$\alpha_{k\ell:\boldsymbol{v}|p}(\omega) := \operatorname{Mod}\left(f_{k\ell:\boldsymbol{v}|p}(\omega)\right) = \sqrt{c^2_{k\ell:\boldsymbol{v}|p}(\omega) + q^2_{k\ell:\boldsymbol{v}|p}(\omega)}, \tag{2.5c}$$

$$\phi_{k\ell:\boldsymbol{v}|p}(\omega) := \operatorname{Arg}\left(f_{k\ell:\boldsymbol{v}|p}(\omega)\right) \in (-\pi, \pi]. \tag{2.5d}$$

Remark 2.3. The sums occurring in eqs. (2.5a) and (2.5b) follows from eq. (2.4b). Equation (2.4a) gives $c_{k\ell:\boldsymbol{v}|p}(\omega) = c_{\ell k|\breve{\boldsymbol{v}}}(\omega)$, $q_{k\ell:\boldsymbol{v}|p}(\omega) = -q_{\ell k|\breve{\boldsymbol{v}}}(\omega)$, $\alpha_{k\ell:\boldsymbol{v}|p}(\omega) = \alpha_{\ell k|\breve{\boldsymbol{v}}}(\omega)$ and $\phi_{k\ell:\boldsymbol{v}|p}(\omega) = -\phi_{\ell k|\breve{\boldsymbol{v}}}(\omega)$.

Remark 2.4. For Gaussian distributions, the local Gaussian correlations will always be equal to the ordinary (global) correlations,[2] and the local Gaussian constructions in definitions 2.1 and 2.3 will thus coincide with the ordinary (global) versions for multivariate Gaussian time series. A comparison of the local and global estimates in the same plot is thus of interest when a given sample is considered, since this could detect nonlinear interactions of the time series under investigation.

Remark 2.5. It is possible to define a local Gaussian anaologue of the squared coherence mentioned in section 1 by replacing the ordinary cross- and auto-spectra with the corresponding local Gaussian versions, i.e. the object of interest would be $\mathcal{Q}_{k\ell:\boldsymbol{v}|p}(\omega) := f_{k\ell:\boldsymbol{v}|p}(\omega) f_{\ell k:\boldsymbol{v}|p}(\omega) / f_{kk:\boldsymbol{v}|p}(\omega) f_{\ell\ell:\boldsymbol{v}|p}(\omega)$. This approach was investigated in the preparation of this paper, but it has not been included here since $\mathcal{Q}_{k\ell:\boldsymbol{v}|p}(\omega)$ in general lacked the nice properties known from the ordinary global case. In particular, the local Gaussian auto-spectra $f_{kk:\boldsymbol{v}|p}(\omega)$ and $f_{\ell\ell:\boldsymbol{v}|p}(\omega)$ will in general be complex valued functions, so an inspection of $\mathcal{Q}_{k\ell:\boldsymbol{v}|p}(\omega)$ must thus be based on plots of its real and imaginary parts (or its amplitude and phase). Moreover, these plots did more often than not turn out to be rather hard to investigate, since the estimates of $f_{kk:\boldsymbol{v}|p}(\omega)$ and $f_{\ell\ell:\boldsymbol{v}|p}(\omega)$ (for some distributions and some frequencies ω) gave values very close to zero in the denominator.

[2] This is due to the way the local Gaussian correlation is defined, see Tjøstheim and Hufthammer (2013) for details.

2.4 Estimation

The estimation of the local Gaussian cross-spectrum $f_{k\ell:\boldsymbol{v}|p}(h)$ from section 2.2 follows the same setup that was used in (Jordanger and Tjøstheim, 2017, section 2.3) for the estimation of the local Gaussian auto-spectrum, with the obvious difference that some extra indices are needed in the present case. The estimation of the related spectra $c_{k\ell:\boldsymbol{v}|p}(h)$, $q_{k\ell:\boldsymbol{v}|p}(h)$, $\alpha_{k\ell:\boldsymbol{v}|p}(h)$ and $\phi_{k\ell:\boldsymbol{v}|p}(h)$ from section 2.3 is then obtained from the estimate of $f_{k\ell:\boldsymbol{v}|p}(h)$ in an obvious manner.

Definition 2.4. *For a sample $\{\boldsymbol{y}_t = (y_{1,t}, \ldots, y_{d,t})\}_{t=1}^n$ of size n from a multivariate time series, an m-truncated estimate $\widehat{f}^m_{k\ell:\boldsymbol{v}|p}(\omega)$ of $f_{k\ell:\boldsymbol{v}|p}(\omega)$ is constructed by means of the following procedure.*

(a) Use the univariate marginals $\{y_{k,t}\}_{t=1}^n$ and $\{y_{\ell,t}\}_{t=1}^n$ to find estimates $\widehat{G}_{k:n}$ and $\widehat{G}_{\ell:n}$ of the corresponding marginal cumulative distribution functions, and compute from this the pseudo-normalised observations $\left\{\widehat{z}_{k,t} \coloneqq \Phi^{-1}\left(\widehat{G}_{k:n}(y_{k,t})\right)\right\}_{t=1}^n$ and $\left\{\widehat{z}_{\ell,t} \coloneqq \Phi^{-1}\left(\widehat{G}_{\ell:n}(y_{\ell,t})\right)\right\}_{t=1}^n$.

(b) Create the lag h pseudo-normalised pairs $\left\{(\widehat{z}_{k,t+h}, \widehat{z}_{\ell,t})\right\}_{t=1}^{n-h}$ for $h = 0, \ldots, m$, and estimate for the point $\boldsymbol{v} = (v_1, v_2)$ the local Gaussian cross-correlations $\left\{\widehat{\rho}_{k\ell:\boldsymbol{v}|p}(h|\boldsymbol{b}_h)\right\}_{h=0}^m$, where the $\{\boldsymbol{b}_h\}_{h=0}^m$ is the bandwidths that are used for the different lags.

(c) Create the lag h pseudo-normalised pairs $\left\{(\widehat{z}_{\ell,t+h}, \widehat{z}_{k,t})\right\}_{t=1}^{n-h}$ for $h = 1, \ldots, m$, and estimate for the diagonally reflected point $\breve{\boldsymbol{v}} = (v_2, v_1)$ the local Gaussian cross-correlations $\left\{\widehat{\rho}_{\ell k:\breve{\boldsymbol{v}}|p}(h|\boldsymbol{b}_h)\right\}_{h=0}^m$.

(d) Adjust eq. (2.4b) from lemma 2.2(b) with some lag-window function $\lambda_m(h)$ to get the estimate

$$\widehat{f}^m_{k\ell:\boldsymbol{v}|p}(\omega) \coloneqq \widehat{\rho}_{k\ell:\boldsymbol{v}|p}(0) + \sum_{h=1}^m \lambda_m(h) \cdot \widehat{\rho}_{\ell k:\breve{\boldsymbol{v}}|p}(h) \cdot e^{+2\pi i\omega h} + \sum_{h=1}^m \lambda_m(h) \cdot \widehat{\rho}_{k\ell:\boldsymbol{v}|p}(h) \cdot e^{-2\pi i\omega h}, \tag{2.6}$$

where the $\{\boldsymbol{b}_h\}_{h=0}^m$ has been suppressed from the notation in order to get a more compact formula.

Definition 2.5. *For a multivariate sample $\{\boldsymbol{y}_t\}_{t=1}^n$ of size n, as described in definition 2.4, the m-truncated estimates of the local Gaussian versions of the cospectrum, quadrature spectrum, amplitude spectrum and phase spectrum is given by*

$$c^m_{k\ell:\boldsymbol{v}|p}(\omega) \coloneqq \operatorname{Re}\left(f^m_{k\ell:\boldsymbol{v}|p}(\omega)\right) = \widehat{\rho}_{k\ell:\boldsymbol{v}|p}(0) + \sum_{h=1}^\infty \cos(2\pi\omega h)\left[\widehat{\rho}_{k\ell:\boldsymbol{v}|p}(h) + \widehat{\rho}_{k\ell:\breve{\boldsymbol{v}}|p}(h)\right], \tag{2.7a}$$

$$q^m_{k\ell:\boldsymbol{v}|p}(\omega) \coloneqq -\operatorname{Im}\left(f^m_{k\ell:\boldsymbol{v}|p}(\omega)\right) = \sum_{h=1}^\infty \sin(2\pi\omega h)\left[\widehat{\rho}_{k\ell:\boldsymbol{v}|p}(h) - \widehat{\rho}_{k\ell:\breve{\boldsymbol{v}}|p}(h)\right], \tag{2.7b}$$

$$\alpha^m_{k\ell:\boldsymbol{v}|p}(\omega) \coloneqq \operatorname{Mod}\left(f^m_{k\ell:\boldsymbol{v}|p}(\omega)\right) = \sqrt{\left(c^m_{k\ell:\boldsymbol{v}|p}(\omega)\right)^2 + \left(q^m_{k\ell:\boldsymbol{v}|p}(\omega)\right)^2}, \tag{2.7c}$$

$$\phi^m_{k\ell:\boldsymbol{v}|p}(\omega) \coloneqq \operatorname{Arg}\left(f^m_{k\ell:\boldsymbol{v}|p}(\omega)\right) \in (-\pi, \pi]. \tag{2.7d}$$

Remark 2.6. The comments in (Jordanger and Tjøstheim, 2017, remarks 2.5 to 2.8) holds for the present case too. In particular, the estimated marginal cumulative distributions $\widehat{G}_{k:n}$ and $\widehat{G}_{\ell:n}$ from definition 2.4(a) can either be based on the (rescaled) empirical cumulative distribution

functions or they could be built upon a logspline technique like the one implemented in Otneim and Tjøstheim (2016). Furthermore, for the asymptotic investigation, the arguments in (Otneim and Tjøstheim, 2016, Section 3) reveals that the pseudo-normalisation of the marginals does not affect the final convergence rates, which (as was done in (Jordanger and Tjøstheim, 2017)) implies that the present theoretical analysis can ignore the distinction between the original observations and the pseudo-normalised observations.

2.5 Asymptotic theory for $\widehat{f}^{m}_{k\ell:\boldsymbol{v}|p}(\omega)$

The asymptotic theory for the local Gaussian cross-spectrum $\widehat{f}^{m}_{k\ell:\boldsymbol{v}|p}(\omega)$ follows from a few minor adjustments of the asymptotic theory that was developed for the local Gaussian auto-spectra. As in (Jordanger and Tjøstheim, 2017, section 2.4), the assumptions and results will be stated for the original observations instead of the pseudo-normalised observations, since this makes the analysis easier and since the final convergence rates are unaffected by this distinction, see remark 2.6 for details.

2.5.1 Some definitions and an assumption for $\boldsymbol{Y}_t$

As for the univariate case in (Jordanger and Tjøstheim, 2017), the assumptions to be imposed on the k and ℓ components of the multivariate times series $\{\boldsymbol{Y}_t\}_{t\in\mathbb{Z}}$ need to be phrased relative to the bivariate pairs that can be created as different combinations of elements from the univariate marginals $\{Y_{k,t}\}_{t\in\mathbb{Z}}$ and $\{Y_{\ell,t}\}_{t\in\mathbb{Z}}$. Note that the *folding property* from item (d) of definition 2.4 implies that it is sufficient to formulate the assumption based on non-negative values of the lag h.

Definition 2.6. *For a strictly stationary multivariate time series $\{\boldsymbol{Y}_t\}_{t\in\mathbb{Z}}$, with $\boldsymbol{Y}_t = \big(Y_{1,t},\ldots,Y_{d,t}\big)$, and for a selected pair of indices k and ℓ, define the following bivariate pairs from the univariate marginals $\{Y_{k,t}\}_{t\in\mathbb{Z}}$ and $\{Y_{\ell,t}\}_{t\in\mathbb{Z}}$.*

$$\boldsymbol{Y}_{k\ell:h:t} \coloneqq \big[Y_{k,t+h}, Y_{\ell,t}\big]', \qquad h \geq 0, \tag{2.8a}$$

$$\boldsymbol{Y}_{\ell k:h:t} \coloneqq \big[Y_{\ell,t+h}, Y_{k,t}\big]', \qquad h \geq 1, \tag{2.8b}$$

and let $g_{k\ell:h}(\boldsymbol{y}_{k\ell:h})$ and $g_{\ell k:h}(\boldsymbol{y}_{\ell k:h})$ denote the respective probability density functions.

The basic idea for the construction of $f_{k\ell:\boldsymbol{v}|p}(\omega)$ is that a point $\boldsymbol{v} = (v_1, v_2)$ should be selected at which for all h the density functions $g_{k\ell:h}(\boldsymbol{y}_{k\ell:h})$ of $\boldsymbol{Y}_{k\ell:h:t}$ will be approximated by $\psi_p\big(\boldsymbol{y}_{k\ell:h}; \boldsymbol{\theta}_{\boldsymbol{v}|k\ell:h|p}\big)$, where ψ_p is one of the bivariate Gaussian density functions from eq. (2.1). The correlation-parameter from the approximating Gaussian density function will be denoted $\rho_{k\ell:\boldsymbol{v}|p}(h)$, and it will be referred to as the local Gaussian lag h cross-correlation of $Y_{k,t+h}$ and $Y_{\ell,t}$ (in that order) at the point $\boldsymbol{v}$.

This local investigation requires a bandwidth vector $\boldsymbol{b} = (b_1, b_2)$ and a kernel function $K(\boldsymbol{w})$, which is used to define $K_{k\ell:h:\boldsymbol{b}}(\boldsymbol{y}_{k\ell:h} - \boldsymbol{v}) \coloneqq \frac{1}{b_1 b_2} K\Big(\frac{y_{k,h}-v_1}{b_1}, \frac{y_{\ell,0}-v_2}{b_2}\Big)$, which in turn is used in

$$q_{\boldsymbol{v}|k\ell:h:\boldsymbol{b}|p} \coloneqq \int_{\mathbb{R}^2} K_{k\ell:h:\boldsymbol{b}}(\boldsymbol{y}_{k\ell:h} - \boldsymbol{v}) \left[\psi_p\big(\boldsymbol{y}_{k\ell:h}; \boldsymbol{\theta}_{\boldsymbol{v}|k\ell:h|p}\big) - \log \psi_p\big(\boldsymbol{y}_{k\ell:h}; \boldsymbol{\theta}_{\boldsymbol{v}|k\ell:h|p}\big)\, g_{k\ell:h}(\boldsymbol{y}_{k\ell:h})\right] \mathrm{d}\boldsymbol{y}_{k\ell:h}, \tag{2.9}$$

a minimiser of which should satisfy the vector equation

$$\int_{\mathbb{R}^2} K_{k\ell:h:\boldsymbol{b}}(\boldsymbol{y}_{k\ell:h} - \boldsymbol{v})\, \boldsymbol{u}_{k\ell:h|p}\big(\boldsymbol{y}_{k\ell:h}; \boldsymbol{\theta}_{\boldsymbol{v}|k\ell:h|p}\big) \left[g_{k\ell:h}(\boldsymbol{y}_{k\ell:h}) - \psi_p\big(\boldsymbol{y}_{k\ell:h}; \boldsymbol{\theta}_{\boldsymbol{v}|k\ell:h|p}\big)\right] \mathrm{d}\boldsymbol{y}_{k\ell:h} = \mathbf{0}, \tag{2.10}$$

where $\boldsymbol{u}_{k\ell:h|p}\big(\boldsymbol{y}_{k\ell:h}; \boldsymbol{\theta}_{\boldsymbol{v}|k\ell:h|p}\big) \coloneqq \boldsymbol{\nabla}_{k\ell:h|p} \log \psi_p\big(\boldsymbol{y}_{k\ell:h}; \boldsymbol{\theta}_{\boldsymbol{v}|k\ell:h|p}\big)$ is the score function of $\psi_p\big(\boldsymbol{y}_{k\ell:h}; \boldsymbol{\theta}_{\boldsymbol{v}|k\ell:h|p}\big)$ (with $\boldsymbol{\nabla}_{k\ell:h|p} \coloneqq \partial/\partial\boldsymbol{\theta}_{\boldsymbol{v}|k\ell:h|p}$). Under the assumption that there is a bandwidth $\boldsymbol{b}_{k\ell:h:0}$ such that there exists a minimiser $\boldsymbol{\theta}_{\boldsymbol{v}|k\ell:h:\boldsymbol{b}|p}$ of eq. (2.9) which satisfies eq. (2.10) for any $\boldsymbol{b}$ with $\mathbf{0} < \boldsymbol{b} < \boldsymbol{b}_{k\ell:h:0}$, this $\boldsymbol{\theta}_{\boldsymbol{v}|k\ell:h:\boldsymbol{b}|p}$ will be referred to as the population value for the given bandwidth $\boldsymbol{b}$.

Remark 2.7. This approach was introduced in a more general context in Hjort and Jones (1996), where it was used to define a local approach to density estimation, and the new idea in (Tjøstheim and Hufthammer, 2013) was to focus upon the estimated local Gaussian parameters $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|k\ell:h|p}$ (instead of the estimated densities). The asymptotic properties of the estimated parameters was investigated in (Tjøstheim and Hufthammer, 2013) by the help of the Klimko-Nelson approach[3] and a suitably defined local penalty function $Q_{\boldsymbol{v}|k\ell:h:n|p}\big(\boldsymbol{\theta}_{\boldsymbol{v}|k\ell:h|p}\big)$ (see eq. (B.1) in appendix B.1).

The assumptions to be imposed on Y_t is related to the estimation of eq. (2.10), and thus requires a few additional definitions.

Definition 2.7. *For $\psi_p\big(\boldsymbol{y}_{k\ell:h}; \boldsymbol{\theta}_p\big)$ the local Gaussian density used when approximating $g_{k\ell:h}(\boldsymbol{y}_{k\ell:h})$ at the point $\boldsymbol{v} = (v_1, v_2)$, and for $\boldsymbol{\theta}_{\boldsymbol{v}|k\ell:h:\boldsymbol{b}|p}$ the population value that minimises the penalty function $q_{\boldsymbol{v}|k\ell:h:\boldsymbol{b}|p}$ from eq. (2.9), define for all $h \in \mathbb{N}$ and all $q \in \{1, \ldots, p\}$*

$$u_{k\ell:h:q:\boldsymbol{b}}(\boldsymbol{w}) \coloneqq \frac{\partial}{\partial\theta_{p:q}} \log\left(\psi_p\big(\boldsymbol{y}_{k\ell:h}; \boldsymbol{\theta}_p\big)\right)\Big|_{\left(\boldsymbol{y}_{k\ell:h}; \boldsymbol{\theta}_p\right) = \left(\boldsymbol{w}; \boldsymbol{\theta}_{\boldsymbol{v}|k\ell:h:\boldsymbol{b}|p}\right)}, \tag{2.11}$$

where $\partial/\partial\theta_{p:q}$ is the q^{th} partial derivative (with respect to $\boldsymbol{\theta}_p$).

The following requirements on the kernel function are identical to those given in (Jordanger and Tjøstheim, 2017, definition B.9).

[3] The Klimko-Nelson approach (see Klimko and Nelson (1978)) 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. The interested reader can consult (Jordanger and Tjøstheim, 2017, appendix B.1) for a more detailed presentation of the Klimko-Nelson approach when a local penalty-function is used.

Definition 2.8. *From a bivariate, non-negative, and bounded kernel function $K(\boldsymbol{w})$, that satisfies*

$$\int_{\mathbb{R}^2} K(w_1, w_2)\, \mathrm{d}w_1 \mathrm{d}w_2 = 1, \tag{2.12a}$$

$$\mathcal{K}_{1:k}(w_2) \coloneqq \int_{\mathbb{R}^1} K(w_1, w_2)\, w_1^k\, \mathrm{d}w_1 \qquad \textit{is bounded for } k \in \{0,1,2\}, \tag{2.12b}$$

$$\mathcal{K}_{2:\ell}(w_1) \coloneqq \int_{\mathbb{R}^1} K(w_1, w_2)\, w_2^\ell\, \mathrm{d}w_2 \qquad \textit{is bounded for } \ell \in \{0,1,2\}, \tag{2.12c}$$

$$\int_{\mathbb{R}^2} K(w_1, w_2)\, |w_1^k w_2^\ell|\, \mathrm{d}w_1 \mathrm{d}w_2 < \infty, \qquad k, \ell \geq 0 \textit{ and } k + \ell \leq 2 \cdot \lceil \nu \rceil\,, \tag{2.12d}$$

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

$$K_{h:\boldsymbol{b}}(\boldsymbol{y}_{k\ell:h} - \boldsymbol{v}) \coloneqq \frac{1}{b_1 b_2} K\left(\frac{y_h - v_1}{b_1}, \frac{y_0 - v_2}{b_2}\right). \tag{2.13}$$

Definition 2.9. *Based on $\boldsymbol{Y}_{k\ell:h:t}$, $u_{k\ell:h:q:\boldsymbol{b}}(\boldsymbol{w})$ and $K_{h:\boldsymbol{b}}(\boldsymbol{y}_{k\ell:h} - \boldsymbol{v})$, define the new bivariate random variables $X^{n|\boldsymbol{v}}_{k\ell:h:q:t}$ as follows,*

$$X^{n|\boldsymbol{v}}_{k\ell:h:q:t} \coloneqq \sqrt{b_1 b_2} K_{h:\boldsymbol{b}}(\boldsymbol{Y}_{k\ell:h:t} - \boldsymbol{v})\, u_{k\ell:h:q:\boldsymbol{b}}(\boldsymbol{Y}_{k\ell:h:t})\,. \tag{2.14}$$

Note that a product of the random variables $X^{n|\boldsymbol{v}}_{k\ell:h:q:t}$ and $X^{n|\boldsymbol{v}}_{k\ell:i:r:s}$ will be a function of $Y_{k,t+h}$, $Y_{\ell,t}$, $Y_{k,s+i}$ and $Y_{\ell,s}$, which depending on the configuration of the indices h, i, s, t will be either a bivariate, trivariate or tetravariate function. The expectation of the product $X^{n|\boldsymbol{v}}_{k\ell:h:q:t} \cdot X^{n|\boldsymbol{v}}_{k\ell:i:r:s}$ will thus (depending on these indices) either require a bivariate, trivariate or tetravariate density function.

Assumption 2.1. The multivariate process $\{\boldsymbol{Y}_t\}_{t\in\mathbb{Z}}$ will be assumed to satisfy the following properties, with $\boldsymbol{v} = (v_1, v_2)$ (in item (d)) the point at which $\widehat{f}^m_{k\ell:\boldsymbol{v}|p}(\omega)$, the estimate of $f_{k\ell:\boldsymbol{v}|p}(\omega)$, is to be computed.

(a) $\{\boldsymbol{Y}_t\}_{t\in\mathbb{Z}}$ is strictly stationary.
(b) $\{\boldsymbol{Y}_t\}_{t\in\mathbb{Z}}$ is strongly mixing, with mixing coefficient $\alpha(j)$ satisfying

$$\sum_{j=1}^{\infty} j^a \left[\alpha(j)\right]^{1-2/\nu} < \infty \qquad \text{for some } \nu > 2 \text{ and } a > 1 - 2/\nu. \tag{2.15}$$

(c) $\mathrm{Var}(||\boldsymbol{Y}_t||^2) < \infty$, where $||\cdot||$ is the Euclidean norm.

The bivariate density functions $g_{k\ell:h}(\boldsymbol{y}_{k\ell:h})$ and $g_{\ell k:h}(\boldsymbol{y}_{\ell k:h})$, corresponding to the lag h pairs introduced in eq. (2.8), must satisfy the following requirements for a given point $\boldsymbol{v} = (v_1, v_2)$.

(d) $g_{k\ell:h}(\boldsymbol{y}_{k\ell:h})$ is differentiable at $\boldsymbol{v}$, such that Taylor's theorem can be used to write $g_{k\ell:h}(\boldsymbol{y}_{k\ell:h})$ as

$$g_{k\ell:h}(\boldsymbol{y}_{k\ell:h}) = g_h(\boldsymbol{v}) + \boldsymbol{\mathfrak{g}}_h(\boldsymbol{v})' \left[\boldsymbol{y}_{k\ell:h} - \boldsymbol{v}\right] + \boldsymbol{\mathfrak{R}}_h(\boldsymbol{y}_{k\ell:h})' \left[\boldsymbol{y}_{k\ell:h} - \boldsymbol{v}\right], \tag{2.16}$$
$$\text{where } \boldsymbol{\mathfrak{g}}_h(\boldsymbol{v}) = \left[\frac{\partial}{\partial y_h} g_{k\ell:h}(\boldsymbol{y}_{k\ell:h}) \Big|_{\boldsymbol{y}_{k\ell:h}=\boldsymbol{v}} , \frac{\partial}{\partial y_0} g_{k\ell:h}(\boldsymbol{y}_{k\ell:h}) \Big|_{\boldsymbol{y}_{k\ell:h}=\boldsymbol{v}} \right]'$$
$$\text{and } \lim_{\boldsymbol{y}_{k\ell:h} \longrightarrow \boldsymbol{v}} \boldsymbol{\mathfrak{R}}_h(\boldsymbol{y}_{k\ell:h}),$$

with the same requirement for $g_{\ell k:h}(\boldsymbol{y}_{\ell k:h})$ at the diagonally reflected point $\breve{\boldsymbol{v}} = (v_2, v_1)$.

(e) There exists a bandwidth $\boldsymbol{b}_{k\ell:h:0}$ such that there for every $\mathbf{0} < \boldsymbol{b} < \boldsymbol{b}_{k\ell:h:0}$ is a unique minimiser $\boldsymbol{\theta}_{\boldsymbol{v}|k\ell:h:\boldsymbol{b}|p}$ of the penalty function $q_{\boldsymbol{v}|k\ell:h:\boldsymbol{b}|p}$ from eq. (2.9).

(f) The collection of bandwidths $\{\boldsymbol{b}_{k\ell:h:0}\}_{h\in\mathbb{Z}}$ has a positive infimum, i.e. there exists a $\boldsymbol{b}_{k\ell:0}$ such that[4]

$$\mathbf{0} < \boldsymbol{b}_{k\ell:0} \coloneqq \inf_{h\in\mathbb{Z}} \boldsymbol{b}_{k\ell:h:0}, \tag{2.17}$$

which implies that this $\boldsymbol{b}_{k\ell:0}$ can be used simultaneously for all the lags.

(g) For $X^{n|\boldsymbol{v}}_{k\ell:h:q:t}$ from definition 2.9, the related bivariate, trivariate and tetravariate density functions must be such that the expectations $\mathrm{E}\left[X^{n|\boldsymbol{v}}_{k\ell:h:q:t}\right]$, $\mathrm{E}\left[\left|X^{n|\boldsymbol{v}}_{k\ell:h:q:t}\right|^{\nu}\right]$ and $\mathrm{E}\left[X^{n|\boldsymbol{v}}_{k\ell:h:q:t} \cdot X^{n|\boldsymbol{v}}_{k\ell:i:r:s}\right]$ all are finite.

Remark 2.8. The present assumption 2.1 is in essence identical to (Jordanger and Tjøstheim, 2017, assumption 2.1) with some extra indices, so the remarks from (Jordanger and Tjøstheim, 2017) is of intereset here too. In particular, the α-mixing requirement in item (b) implies that $Y_{k,t+h}$ and $Y_{\ell,t}$ will be asymptotically independent as $h \to \infty$, i.e. the bivariate density functions $g_{k\ell:h}(\boldsymbol{y}_{k\ell:h})$ 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 $\boldsymbol{b}_{k\ell:0}$ that works for all h. Moreover, the finiteness assumptions in item (g) are always trivially satisfied if the required density-functions are finite.

2.5.2 An assumption for $\left(Y_{k,t}, Y_{\ell,t}\right)$ and the score function $\boldsymbol{u}_p(\boldsymbol{w};\boldsymbol{\theta}_p)$ of $\psi_p(\boldsymbol{w};\boldsymbol{\theta}_p)$

The following assumption is in essence identical to (Jordanger and Tjøstheim, 2017, assumption 2.2), which was included due to the need for the asymptotic results from Tjøstheim and Hufthammer (2013) to be applied for all the different lags h.

Assumption 2.2. The collection of local Gaussian parameters $\left\{\boldsymbol{\theta}_{\boldsymbol{v}|k\ell:h|p}\right\}$ at the point $\boldsymbol{v}$ for the bivariate probability density functions $g_{k\ell:h}(\boldsymbol{y}_{k\ell:h})$, must all be such that

(a) $\boldsymbol{u}_p(\boldsymbol{v};\boldsymbol{\theta}_{\boldsymbol{v}|k\ell:h|p}) \neq \mathbf{0}$ for all finite h.

(b) $\lim \boldsymbol{u}_p(\boldsymbol{v};\boldsymbol{\theta}_{\boldsymbol{v}|k\ell:h|p}) \neq \mathbf{0}$.

Note that an inspection of the p equations in $\boldsymbol{u}_p(\boldsymbol{w};\boldsymbol{\theta}_p) = \mathbf{0}$ can be used to identify when items (a) and (b) of assumption 2.2 might fail, cf. the discussion in (Jordanger and Tjøstheim, 2017, section 2.4.2) for further details.

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

2.5.3 Assumptions for n, m and $\boldsymbol{b} = (b_1, b_2)$

The following assumption is identical to (Jordanger and Tjøstheim, 2017, assumption 2.3). The internal consistency of it was verified in (Jordanger and Tjøstheim, 2017, lemma C.3).

Assumption 2.3. Let $m \coloneqq m_n \to \infty$ be a sequence of integers denoting the number of lags to include, and let $\boldsymbol{b} \coloneqq \boldsymbol{b}_n \to \mathbf{0}^+$ be the bandwidths used when estimating the local Gaussian correlations for the lags $h = 1, \dots, m$ (based on n observations). Let b_1 and b_2 refer to the two components of $\boldsymbol{b}$, and let α, ν and a be as introduced in assumption 2.1(b). Let $s \coloneqq s_n \to \infty$ be a sequence of integers such that $s = o\Big(\sqrt{nb_1b_2/m}\Big)$, and let τ be a positive constant. The following requirements must be satisfied for these entities.[5]

(a) $\log n/n(b_1b_2)^5 \longrightarrow 0$, (only required for the case $p = 5$).
(b) $nb_1b_2/m \longrightarrow \infty$.
(c) $m^\delta(b_1 \vee b_2) \longrightarrow 0$, where $\delta = 2 \vee \frac{\nu(a+1)}{\nu(a-1)-2}$.
(d) $\sqrt{nm/b_1b_2} \cdot s^\tau \cdot \alpha(s-m+1) \longrightarrow \infty$.
(e) $m = o\big((nb_1b_2)^{\tau/(2+5\tau)-\lambda}\big)$, for some $\lambda \in (0, \tau/(2+5\tau))$.
(f) $m = o(s)$.

2.5.4 Convergence theorems for $\widehat{f}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega)$, $\widehat{\alpha}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega)$ and $\widehat{\phi}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega)$

See appendix A for the proofs of the theorems stated below.

Theorem 2.10. *The estimate $\widehat{f}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega) = \widehat{c}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega) - i \cdot \widehat{q}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega)$ of the local Gaussian cross-spectrum $f_{k\ell:\boldsymbol{v}|p}(\omega) = c_{k\ell:\boldsymbol{v}|p}(\omega) - i \cdot q_{k\ell:\boldsymbol{v}|p}(\omega)$, will under assumptions 2.1 to 2.3 satisfy*

$$\sqrt{n(b_1b_2)^{(p+1)/2}/m} \cdot \left(\begin{bmatrix} \widehat{c}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega) \\ \widehat{q}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega) \end{bmatrix} - \begin{bmatrix} c_{k\ell:\boldsymbol{v}|p}(\omega) \\ q_{k\ell:\boldsymbol{v}|p}(\omega) \end{bmatrix} \right) \xrightarrow{d} \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{c|k\ell:\boldsymbol{v}|p}(\omega) & 0 \\ 0 & \sigma^2_{q|k\ell:\boldsymbol{v}|p}(\omega) \end{bmatrix} \right), \tag{2.18}$$

when $\omega \notin \frac{1}{2} \cdot \mathbb{Z} = \big\{\dots, -1, -\frac{1}{2}, 0, \frac{1}{2}, 1, \dots\big\}$, where the variances $\sigma^2_{c|k\ell:\boldsymbol{v}|p}(\omega)$ and $\sigma^2_{q|k\ell:\boldsymbol{v}|p}(\omega)$ are given by

$$\sigma^2_{c|k\ell:\boldsymbol{v}|k\ell|p}(\omega) = \lim_{m\to\infty} \frac{1}{m} \left(\tilde{\sigma}^2_{\boldsymbol{v}|k\ell|p}(0) + \sum_{h=1}^{m} \lambda^2_m(h) \cdot \cos^2(2\pi\omega h) \cdot \big\{ \tilde{\sigma}^2_{\boldsymbol{v}|k\ell|p}(h) + \tilde{\sigma}^2_{\check{\boldsymbol{v}}|\ell k|p}(h) \big\} \right) \tag{2.19a}$$

$$\sigma^2_{q|k\ell:\boldsymbol{v}|k\ell|p}(\omega) = \lim_{m\to\infty} \frac{1}{m} \left(\sum_{h=1}^{m} \lambda^2_m(h) \cdot \sin^2(2\pi\omega h) \cdot \big\{ \tilde{\sigma}^2_{\boldsymbol{v}|k\ell|p}(h) + \tilde{\sigma}^2_{\check{\boldsymbol{v}}|\ell k|p}(h) \big\} \right), \tag{2.19b}$$

with $\tilde{\sigma}^2_{\boldsymbol{v}|k\ell|p}(h)$ and $\tilde{\sigma}^2_{\check{\boldsymbol{v}}|\ell k|p}(h)$ the asymptotic variances related to the estimates $\widehat{\rho}_{k\ell:\boldsymbol{v}|p}(h)$ and $\widehat{\rho}_{\ell k:\check{\boldsymbol{v}}|p}(h)$, see theorem B.1 for the details.

The local Gaussian quadrature spectrum is identical to zero when $\omega \in \frac{1}{2} \cdot \mathbb{Z}$, and for those frequencies the following asymptotic result holds under the given assumptions

$$\sqrt{n(b_1b_2)^{(p+1)/2}/m} \cdot \Big(\widehat{f}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega) - f_{k\ell:\boldsymbol{v}|p}(\omega) \Big) \xrightarrow{d} \mathrm{N}\big(0, \sigma^2_{c|k\ell:\boldsymbol{v}|p}(\omega)\big), \qquad \omega \in \tfrac{1}{2} \cdot \mathbb{Z}. \tag{2.20}$$

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

The asymptotic results for the local Gaussian amplitude- and phase-spectra is a direct consequence of theorem 2.10 and (Brockwell and Davis, 1986, proposition 6.4.3, p. 211).

Theorem 2.11. *Under assumptions 2.1 to 2.3, when* $\alpha_{k\ell:\boldsymbol{v}|p}(\omega) > 0$ *and* $\omega \notin \frac{1}{2} \cdot \mathbb{Z}$*, the estimate* $\widehat{\alpha}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega) = \sqrt{\left(\widehat{c}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega)\right)^2 + \left(\widehat{q}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega)\right)^2}$ *satisfies*

$$\sqrt{n(b_1 b_2)^{(p+1)/2}/m} \cdot \left(\widehat{\alpha}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega) - \alpha_{k\ell:\boldsymbol{v}|p}(\omega)\right) \xrightarrow{d} \mathrm{N}(0, \sigma^2_\alpha(\omega)), \tag{2.21}$$

where $\sigma^2_\alpha(\omega)$ *is given relative to* $\sigma^2_{c|k\ell:\boldsymbol{v}|p}(\omega)$ *and* $\sigma^2_{q|k\ell:\boldsymbol{v}|p}(\omega)$ *(from eq.* (A.9)*in theorem 2.10) as*

$$\sigma^2_\alpha = \left(c^2_{k\ell:\boldsymbol{v}|p}(\omega) \cdot \sigma^2_{c|k\ell:\boldsymbol{v}|p}(\omega) + q^2_{k\ell:\boldsymbol{v}|p}(\omega) \cdot \sigma^2_{q|k\ell:\boldsymbol{v}|p}(\omega)\right) / \alpha^2_{k\ell:\boldsymbol{v}|p}(\omega). \tag{2.22}$$

Theorem 2.12. *Under assumptions 2.1 to 2.3, when* $\alpha_{k\ell:\boldsymbol{v}|p}(\omega) > 0$ *and* $\omega \notin \frac{1}{2} \cdot \mathbb{Z}$*, the estimate* $\widehat{\phi}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega) = \operatorname{args}\left(\widehat{c}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega) - i \cdot \widehat{q}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega)\right)$ *satisfies*

$$\sqrt{n(b_1 b_2)^{(p+1)/2}/m} \cdot \left(\widehat{\phi}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega) - \phi_{k\ell:\boldsymbol{v}|p}(\omega)\right) \xrightarrow{d} \mathrm{N}\left(0, \sigma^2_\phi(\omega)\right), \tag{2.23}$$

where $\sigma^2_\phi(\omega)$ *is given relative to* $\sigma^2_{c|k\ell:\boldsymbol{v}|p}(\omega)$ *and* $\sigma^2_{q|k\ell:\boldsymbol{v}|p}(\omega)$ *(from eq.* (A.9)*in theorem 2.10) as*

$$\sigma^2_\phi(\omega) = \left(q^2_{k\ell:\boldsymbol{v}|p}(\omega) \cdot \sigma^2_{c|k\ell:\boldsymbol{v}|p}(\omega) + c^2_{k\ell:\boldsymbol{v}|p}(\omega) \cdot \sigma^2_{q|k\ell:\boldsymbol{v}|p}(\omega)\right) / \alpha^4_{k\ell:\boldsymbol{v}|p}(\omega). \tag{2.24}$$

Remark 2.9. The asymptotic normality results in theorems 2.10 to 2.12 do not necessarily help much if computations of pointwise confidence intervals for the estimated local Gaussian estimates are of interest, since it in practice may be unfeasible to find decent estimates of the variances $\sigma^2_{c|k\ell:\boldsymbol{v}|p}(\omega)$ and $\sigma^2_{q|k\ell:\boldsymbol{v}|p}(\omega)$ that occurs in theorem 2.10. The pointwise confidence intervals will thus later on 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.

3 Examples

This section will investigate if the m-truncated estimates of the local Gaussian cross-spectrum $f_{k\ell:\boldsymbol{v}|p}(\omega)$ might be of interest to consider when multiviarate time series are encountered. Since $\widehat{f}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega)$ is complex-valued, the actual investigation will be based on plots of the corresponding local Gaussian versions of the cospectrum, quadrature spectrum, phase spectrum and the amplitude spectrum.

The present setup is similar to the one used for the investigation of the local Gaussian auto-spectrum, see (Jordanger and Tjøstheim, 2017, section 3), i.e. it will first be checked that the expected result is obtained when the origin of the data is a bivariate Gaussian time series. After this a simulation from a bivariate extension of the *local trigonometric* time series from (Jordanger and Tjøstheim, 2017, section 3.3.2) will be considered, and there will moreover be an example based on real multivariate data together with a simulation from a model based on these data. As in (Jordanger and Tjøstheim, 2017), it will be seen that plots related to the estimated $\widehat{f}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega)$ might be useful as an exploratory tool, i.e. that this approach can detect nonlinear dependencies and periodicities between the variables.

Several parameters must be specified in order for an m-truncated estimate of $f_{k\ell:\boldsymbol{v}|p}(\omega)$ to be computed, and for the examples in the present paper the following values will be used:

1. p, the number of parameters in the local Gaussian approximation. Only the value $p = 5$ will be used in the present examples, since the results based on $p = 1$ in general fails to capture the local structure in a proper manner (see the discussion in (Jordanger and Tjøstheim, 2017, section 3.7) for further details).
2. $\boldsymbol{v} = (v_1, v_2)$, the points to investigate. The present investigation will consider points whose first and second coordinates correspond 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 will be marked as `10%::50%` and so on.
3. ω, the frequencies to investigate. Values between 0 and $\frac{1}{2}$.
4. $\boldsymbol{b} = (b_1, b_2)$, the bandwidth-vector to be used when computing the local Gaussian autocorrelations. The value $\boldsymbol{b} = (.6, .6)$ has been used for all of the cases in this paper.
5. m, the truncation level, i.e. the number of lags to include in the estimate $\widehat{f}^{\,m}_{\boldsymbol{v}|p}(\omega)$. 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.
6. $\lambda_m(h)$, 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.

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

Remark 3.1. This list of parameters is similar to the one used when estimating the local Gaussian auto-spectra in (Jordanger and Tjøstheim, 2017, section 3.1), with the main exception that the value 0.6 is used instead of 0.5 in the bandwidth vector $\boldsymbol{b}$. This adjustment is partially due to the fact that the time series in this paper are sligthly shorter than those used in (Jordanger and Tjøstheim, 2017), i.e. length 1859 versus length 1974. (All the time series have the same length as the one encountered for the real sample.)

Remark 3.2. It was noted in (Jordanger and Tjøstheim, 2017) that it was natural to require that the bandwidth $\boldsymbol{b} = (b_1, b_2)$ should satisfy $b_1 = b_2$ when the local Gaussian autocorrelations $\rho_{kk:\boldsymbol{v}|p}(h)$ should be estimated, since both of the components in the lag h pseudo-normalised pairs originated from the same univariate time series. For the estimation of local Gaussian cross-correlations $\rho_{k\ell:\boldsymbol{v}|p}(h)$, it is the pseudo-normalisation of the marginals that justifies the assumed equality of b_1 and b_2.

Remark 3.3. The *pointwise confidence bands*[6] 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 estimated 90% pointwise confidence bands for $c^m_{k\ell:\boldsymbol{v}|p}(\omega)$, $q^m_{k\ell:\boldsymbol{v}|p}(\omega)$, $\alpha^m_{k\ell:\boldsymbol{v}|p}(\omega)$, and $\phi^m_{k\ell:\boldsymbol{v}|p}(\omega)$.

[6] The pointwise confidence band gives for each frequency ω a confidence interval for the values of $c^m_{k\ell:\boldsymbol{v}|p}(\omega)$, $q^m_{k\ell:\boldsymbol{v}|p}(\omega)$, $\alpha^m_{k\ell:\boldsymbol{v}|p}(\omega)$, and $\phi^m_{k\ell:\boldsymbol{v}|p}(\omega)$.

Remark 3.4. It will in general be hard to know in advance which input parameters it would be natural to employ for a given time series, and it is thus necessary and recommended to find estimates $\widehat{f}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega)$ for several different parameter combinations and then use an interactive tool to inspect the corresponding plots to see if some interesting features might be present. The `R`-package `localgaussSpec`[7] has been created to take care of both of these parts, i.e. it allows an integrated interactive investigation of the results by means of a `shiny`-application.[8]

Remark 3.5. The scripts that generated the plots presented in this paper is included as a part of the `R`-package `localgaussSpec`, so the interested reader can run these and check out how the plots change when the input parameters in the estimation algorithm are modified. Note that any conclusions based on individual static plots, like those presented in this paper, in general should be considered with some caution, in particular when there is a lack of data-driven methods that can justify the parameter-configuration that was used in the estimation-algorithm. Moreover, it is important to be aware of the fact that small-sample variation might occur not only if an estimate is computed for a time series sample that is too short, but that it also can be small-sample variation that dominates if the point $\boldsymbol{v}$ lies too far out in the periphery, or if the truncation level m and the bandwidth $\boldsymbol{b}$ are not keept within reasonable limits. See the discussion in section 4 for further details.

Remark 3.6. The `R`-package `localgauss`, see Berentsen et al. (2014), was used for the estimation of the local Gaussian auto- and cross-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.

3.1 Some basic simulations

This section will check that the estimates of $f^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega)$ behaves as expected for a few simple simulated bivariate examples, where the underlying models in essence are bivariate extensions of the models encountered in (Jordanger and Tjøstheim, 2017, section 3.3), and it will be seen that a joint inspection of the `Co`-, `Quad`- and `Phase`-plots might be useful as an exploratory tool for nonlinear dependencies in multivariate time series. The `Amplitude`-plots have not been included here since the interesting details (in most cases) already would have been detected by the other plots.

3.1.1 Bivariate Gaussian white noise

It is known from lemma 2.2(a) that the local Gaussian cross-spectrum coincide with the ordinary cross-spectrum when the time series under investigation is Gaussian. The plots in fig. 1 shows the `Co`-, `Quad`- and `Phase`-plots based on 100 independent samples of length 1859 from a bivariate Gaussian distribution with standard normal marginals and correlation 0.35. The left column of fig. 1 shows the situation for a point off the diagonal, whereas the right column

[7] See footnote 1 (page 3) for details about the installation of the `localgaussSpec`-package.
[8] See Chang et al. (2017) for details about `shiny`.

shows the situation for a point at the center of the diagonal, i.e. $v_1 = v_2 = 0$. Note that the global spectra are identical for all the points, i.e. the red components are the same for each row of fig. 1.

In this simple case, where the true values of the local Gaussian versions of the spectra coincides with the ordinary global spectra, it follows that the `Co`-, `Quad`- and `Phase`-spectra (for any truncation level m) respectively should be the constants 0.35, 0 and 0. Figure 1 shows that the red and blue dotted lines, that respectively represents the estimates of the global and local m-truncated spectra,[9] seems to match these true values quite reasonably – and this provides a sanity check of the code that generated these plots. Note that the 90% pointwise confidence interval for the local Gaussian versions (blue ribbons) are wider than those for the ordinary spectra (red ribbons) since the bandwidth used for the estimation of the local Gaussian cross-correlations, in this case $\boldsymbol{b} = (0.6, 0.6)$, reduces the number of observations that effectively contributes to the computation of the local Gaussian spectra.

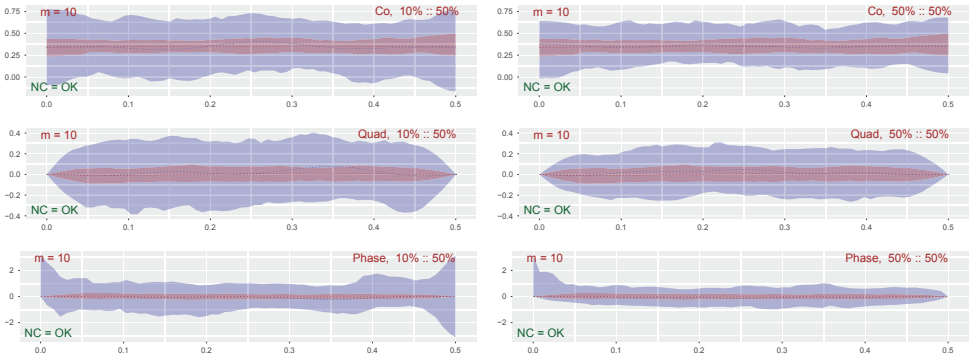

Figure 1: i.i.d. bivariate Gaussian white noise.

3.1.2 Bivariate local trigonometric examples

With the exception of the Gaussian time series, it is not known what the true local Gaussian spectral densities should look like – which makes it hard to know whether or not the `Co`-, `Quad`- and `Phase`-plots looks like they are expected to do. However, it is possible to extend to the bivariate case the strategy that was outlined in (Jordanger and Tjøstheim, 2017, section 3.3.2), i.e. that *bivariate local trigonometric time series* can be constructed for which it, at some designated points $\boldsymbol{v}$, can be given a heuristic argument for the expected shape of the local Gaussian spectra.

The heuristic argument needed for the bivariate case is identical in structure to the one used in the univariate case, and for the present case the reference for the plots later on is based on the following simple bivariate model,

$$Y_{1,t} = \cos(2\pi\alpha t + \phi) + w_t \text{ and } Y_{2,t} = \cos(2\pi\alpha t + \phi + \theta) + w_t, \tag{3.1}$$

[9] The dotted lines represents the medians of the estimated values, whereas the 90% pointwise confidence intervals are based on the 5% and 95% quantiles of these samples.

where w_t is Gaussian white noise with mean zero and standard deviation σ, and where it in addition is such that α and θ are fixed for all the replicates whereas ϕ is drawn uniformly from $[0, 2\pi)$ for each individual replicate. A realisation with $\sigma = 0.75$, $\alpha = 0.302$ and $\theta = \pi/3$ has been used for the `Co-`, `Quad-`, and `Phase`-plots shown in fig. 2, where 100 independent samples of length 1859 were used to get the estimates of the m-truncated spectra and their corresponding 90 pointwise confidence intervals (based on the bandwidth $\boldsymbol{b} = (0.6, 0.6)$). Some useful remarks can be based on this plot, before *the bivariate local trigonometric case* is defined and investigated.

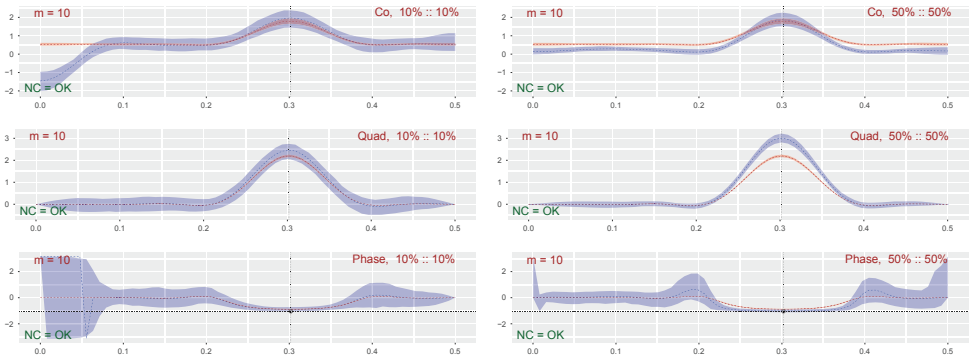

Figure 2: Realisation of eq. (3.1), with $\alpha = 0.302$ and $\theta = \pi/3$.

Remark 3.7. In this particular case, the local Gaussian spectra in fig. 2 shares many similarities with the corresponding global spectra. In particular, the peak of the `Co-` and `Quad`-plots lies both for the local and global spectra at the frequency $\omega = \alpha$ (shown in the plots as a vertical line), and the corresponding `Phase`-plots at this frequency lies quite close to the phase-adjustment $\theta = \pi/3$ (shown as a horizontal line, positioned with an appropriate sign adjustment). This phenomenon is present both for the point at `10%::10%` and the point at `50%::50%`, but it should be noted that this nice match does not hold for all values of σ. In fact, experiments with different values for σ (plots not included in this paper) indicates that the difference between the local and global spectra becomes larger (in particular for the point `10%::10%`) when σ becomes smaller. See (Jordanger and Tjøstheim, 2017, fig. 9) for an example of the results that can occur.

Remark 3.8. The values of the `Co-` and `Quad`-plots (for a given frequency ω) are (for each frequency) related to the corresponding values of the `Amplitude-` and `Phase`-plots by the following simple relations,

$$-\texttt{Quad-plot}/\texttt{Co-plot} = \tan(\texttt{Phase-plot}), \tag{3.2a}$$

$$\texttt{Co-plot} = \texttt{Amplitude-plot} \cdot \cos(\texttt{Phase-plot}), \tag{3.2b}$$

$$-\texttt{Quad-plot} = \texttt{Amplitude-plot} \cdot \sin(\texttt{Phase-plot}), \tag{3.2c}$$

which follows trivially from the way these spectra are defined relative to Cartesian or polar representations of the complex-valued cross-spectra, cf. eq. (1.2) and definition 2.3. For the example investigated in fig. 2, where the `Phase` plot is close to $-\pi/3$ at $\alpha = 0.302$, it thus

follows that the peak for the Quad-spectrum should be approximately $\sqrt{3}$ times larger than the peak of the Co-spectrum.

Remark 3.9. It can be enlightening to compare the Co-, Quad- and Phase-plots in fig. 2 with a plot that shows the underlying estimates upon which the pointwise confidence intervals were based. Such a plot is shown in fig. 3, where the left panel presents the complex-valued estimates of the local Gaussian cross-spectrum at the frequency $\omega = \alpha$, and where medians and quantiles relative to a polar representation, i.e. $z = re^{i\theta}$, have been added to the plot. The center panel shows the same estimated values, but this time the median/quantiles are based on a Cartesian representation $z = x + iy$. These two panels gives a geometrical view to the observations presented in remark 3.8. The third panel of fig. 3 presents a zoomed in version of the estimated values of the local Gaussian cross-spectrum, and it gives a reminder that it in principle is possible to extract more information from these estimates than what has been done so far. A closer inspection of these estimates could e.g. be used to see how much they (for the given m-truncation) deviate from the expected asymptotic distributions that was given in section 2.5.4.

Remark 3.10. The wide pointwise confidence band observed for ω near 0 in the 10%::10%-Phase-plot, is related to the branch-cut that occurs at $-\pi$ in the definition of the phase-spectrum, cf. definition 2.3. The simple algorithm used for the creation of the pointwise confidence intervals has not been tweaked to properly cover the case where the majority of the estimates lies in the second and third quadrants of the complex plane, which implies that the Co- and Quad-plots should be consulted instead when the Phase-plot misbehaves in this manner.

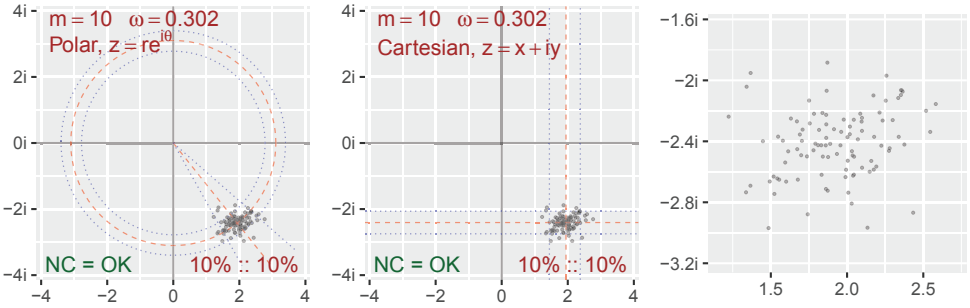

Figure 3: Complex-valued representation of 100 samples of $f_{12;v}(\omega)$ from eq. (3.1), at the peak frequency $\omega = 0.302$. Left panel: Pointwise 90% confidence bands based on polar representation. Center panel: Pointwise 90% confidence bands based on Cartesian representation. Right panel: Zoomed in plot.

The bivariate local trigonometric case: Two bivariate extensions of the artificial *local trigonometric* time series from (Jordanger and Tjøstheim, 2017, section 3.3.2) will now be considered. The parameters in these time series can be selected in such a manner that the global spectra are similar to those encountered for white noise, whereas the local Gaussian spectra contains different peaks at different points. The same heuristic argument that was used in (Jordanger and Tjøstheim, 2017) implies that it is *expected* that the local periodicity at some designated points should resemble those in fig. 2, and samples from these models can thus be

used to check if the `Co-`, `Quad-` and `Phase`-plots can detect these local properties. It will be seen that the answer indeed is affirmative for the designated points for both of the models considered here, see figs. 4 and 6, but it will also be noted (see fig. 5) that the plots can be quite different when other points are investigated.

The algorithm for the bivariate samples $\left(Y_{1,t}, Y_{2,t}\right)$ (defined below) is based on the univariate algorithm from (Jordanger and Tjøstheim, 2017), i.e. the marginals $Y_{1,t}$ and $Y_{2,t}$ will both be generated by the same *local trigonometric formula* that was used in (Jordanger and Tjøstheim, 2017, section 3.3.2), but with the distinction that an additional phase-adjustment is used for the second component. To clarify this, here follows the required modification of the univariate setup from (Jordanger and Tjøstheim, 2017), with the specification of the parameters used in the present computations.

The *bivariate local trigonometric* time series are constructed by the following principle: For a given $r \geq 2$, first select a collection of different base levels $(L_1, \ldots, L_r)$ at the y-axis, and a collection of amplitudes $(A_1, \ldots, A_r)$ and amplitude adjustments $(A'_1, \ldots, A'_r)$. Create a stochastic amplitude $A_i(t)$, for each t, by selecting uniformly a value from the interval spanned by A_i and A'_i. Then select a collection of frequencies $(\alpha_1, \ldots, \alpha_r)$, and two collections of phase-adjustments $(\phi_1, \ldots, \phi_r)$ and $(\theta_1, \ldots, \theta_r)$. Finally, assign a probability p_i to each $i = 1, \ldots, r$, such that $\sum_{i=1}^{r} p_i = 1$. These ingredients enables the definition of the following functions for $i = 1, \ldots, r$,

$$C_{1,i}(t) = L_i + A_i(t) \cdot \cos\left(2\pi\alpha_i t + \phi_i\right), \tag{3.3a}$$

$$C_{2,i}(t) = L_i + A_i(t) \cdot \cos\left(2\pi\alpha_i t + \phi_i + \theta_i\right), \tag{3.3b}$$

from which the stochastic variables $Y_{1,t}$ and $Y_{2,t}$ can be created by means of the probabilities $(p_1, \ldots, p_r)$, i.e. let N_t be a random variable that with probability p_i takes the value i, and define

$$Y_{1,t} \coloneqq \sum_{i=1}^{r} C_{1,i}(t) \cdot \mathbb{1}\{N_t = i\}, \tag{3.4a}$$

$$Y_{2,t} \coloneqq \sum_{i=1}^{r} C_{2,i}(t) \cdot \mathbb{1}\{N_t = i\}. \tag{3.4b}$$

The indicator function $\mathbb{1}\{\}$ ensures that only one of the r functions in the sum contributes for 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 auto- and cross-covariances can can be given as functions of L_i and p_i, from which it then is fairly easy to select a combination of input parameters that returns a $\left(Y_{1,t}, Y_{2,t}\right)$-process that looks like white noise.

The generating models for the two time series presented in this section both have $r = 4$ 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 $(0.267, 0.091, 0.431, 0.270)$.

The probabilities p_i in $(0.05, 0.28, 0.33, 0.33)$ was used to sample[10] which component to include in $Y_{1,t}$ and $Y_{2,t}$.

The distinction between the two models are due to the selection of the additional phase-adjustments θ_i. The model investigated in figs. 4 and 5 have a constant phase adjustment of $\theta = \pi/3$, whereas the model investigated in fig. 6 have individual phase-adjustments given as $(\theta_1, \theta_2, \theta_3, \theta_4) = (\pi/3, \pi/4, 0, \pi/2)$.

To complete the specification of the setup, note that the 90% pointwise confidence intervals in figs. 4 to 6 all are based on 100 independent samples of length 1859 from the above described models, and that the bandwidth $\boldsymbol{b} = (0.6, 0.6)$ was used in the computation of the local Gaussian cross-correlations.

Constant phase adjustment: The case where the phase difference $\theta = \pi/3$ was used for all the θ_i is investigated in figs. 4 and 5. Figure 4 (see page 22) shows the `Co-`, `Quad-` and `Phase`-plots for the three designated points `10%::10%`, `50%::50%` and `90%::90%`, for which the heuristic argumentation implies that the results should look a bit like the situation encountered in fig. 2.

Remark 3.11. The three points investigated in fig. 4 corresponds to the function-components in eq. (3.3) with indices $i = 2, 3, 4$. The point that corresponds to the $i = 1$ component would, due to the combination of the low probability p_1 and the placement of the level L_1, lie too far out in the tail to be properly investigated by the present sample size and truncation point. The $i = 1$ component was included in the example in (Jordanger and Tjøstheim, 2017) in order to show that an exploratory approach based on local Gaussian spectra can fail to detect local signals that are much weaker than the dominating ones.

For the points investigated in fig. 4, it seems to be the case that the local Gaussian part of the `Co-`, `Quad-` and `Phase`-plots together reveal local properties in accordance with the outcome expected from the knowledge of the generating model – and these local structures are not detected by the ordinary global spectra, which in this case (due to the values used for L_i and p_i) are flat. The left column investigates a point at the lower tail of the diagonal, and it can there be observed that both the `Co-` and `Quad`-plots have a peak close to the leftmost α-value – and the value of the corresponding `Phase`-plot for frequencies close to this α-value lies quite close to the phase difference between the first and second component. A similar situation is present for the three plots shown in the right column, where a point at the upper tail of the diagonal are investigated. Moreover, in accordance with the general observation in remark 3.8, the peaks of the `Quad`-plots are higher than those of the `Co`-plots in this case due to the phase-difference θ that was used in the input parameters.

For the center column of fig. 4, which investigates the point at the center of the diagonal, it can be seen that the `Quad`- and `Phase`-plots in addition to the expected α-value also detects the presence of the other α-values. The `Phase`-plot is for this point not that close to the expected value, but that situation changes if the truncation is performed at a higher lag than $m = 10$. The center column thus shows the importance of considering a range of values for the truncation point when such plots are investigated. The additional peaks that are detected in the center column are due to *contamination* from the neighbouring regions.

[10] The 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 5 present plots based on the same sample that was used for fig. 4, but now the investigated points are no longer among the designated ones on the diagonal. The plots in fig. 5 shows that the Co-, Quad- and Phase-plots at the point 10%::90% looks more like the i.i.d. white noise that was encountered in fig. 1, whereas the plots for the two points 10%::50% and 50%::90% does detect the presence of local phenomena. It might not be any obvious interpretation of these plots when seen isolated, but it should at least be noted that the plots for the two points 10%::50% and 50%::90% have troughs for the α-values that corresponds to the first and second coordinates of these points – and this seen in conjunction with the previously investigated points in fig. 4 does support the idea that there are local features in the data that depends on these α-values.

Individual phase adjustment: Figure 6 investigates the same designated points as fig. 4 did, but now the samples have been generated from the model where the second variate used the individual phase-adjustments specified in $(\theta_1, \theta_2, \theta_3, \theta_4) = (\pi/3, \pi/4, 0, \pi/2)$. Horizontal lines have been added to the Phase-plots to show all of these θ_i-values (adjusted to have the correct sign), and for each of the designated points the intersection with the relevant vertical α_i-line has been highlighted to show the expected outcome based on the knowledge of the model.

The Co-, Quad- and Phase-plots in fig. 6 does behave in accordance with what was observed in fig. 4, i.e. the Phase-plots lies close to the expected θ_i-value when the frequency ω is near the corresponding α_i-value, and the height of the corresponding Co- and Quad-peaks are in accordance with the values of the Phase-plots. In particular, the phase-adjustment is $\theta_2 = \pi/4$ for the point 10%::10%, which implies that the Co- and Quad-peaks should rise approximately to the same height above their respective baselines, which seems to be fairly close to the observed result. For the points 50%::50% and 90%::90% the situation is clearer since the respective local frequencies $\theta_3 = 0$ and $\theta_4 = \pi/2$ then implies that only the Co-plot should have a peak for the point 50%::50% and only the Quad-plot should have a peak for the point 90%::90%, again in agreement with the impression based on fig. 6.

Remark 3.12. The examples investigated in figs. 2 to 6 do not satisfy the requirements needed for the asymptotic results (both for the global and local cases) to hold true, in particular the local Gaussian cross-correlations will in these cases not be absolutely summable.[11] Despite this, the examples are still of interest since they shows that an exploratory tool based on the local Gaussian spectra in this case can detect information that is in agreement with what could be expected based on the parameters used in the models for $\left(Y_{1,t}, Y_{2,t}\right)$.

Remark 3.13. The generating model will of course not be known when a real multivariate time series is encountered, so it is important to estimate the local Gaussian cross-spectrum at a wide range of points $\boldsymbol{v}$ in the plane and a wide range of truncation levels m. This is necessary since it can happen, like seen in fig. 5, that there are some points where the result might look like it has been computed based on i.i.d. white noise, whereas other points shows that some local phenomenon could be present.

Remark 3.14. Even when it might not be obvious how to interpret the results shown in the Co-, Quad- and Phase-plots, it should be noted that they can be used as an exploratory tool that can detect nonlinear traits in the observations. Moreover, these plots can also be used to investigate if a model fitted to the data contains elements that can mimic the observed features. The recipe

[11] In this respect the situation is similar to the detection of a pure sinusoidal for the global spectrum.

for this approach would then be to first select a model, then estimate parameters based on the available sample, and finally use the resulting fitted model to generate independent samples of the same length as the sample. Section 3.2 will show an example of this approach, cf. figs. 9 and 12.

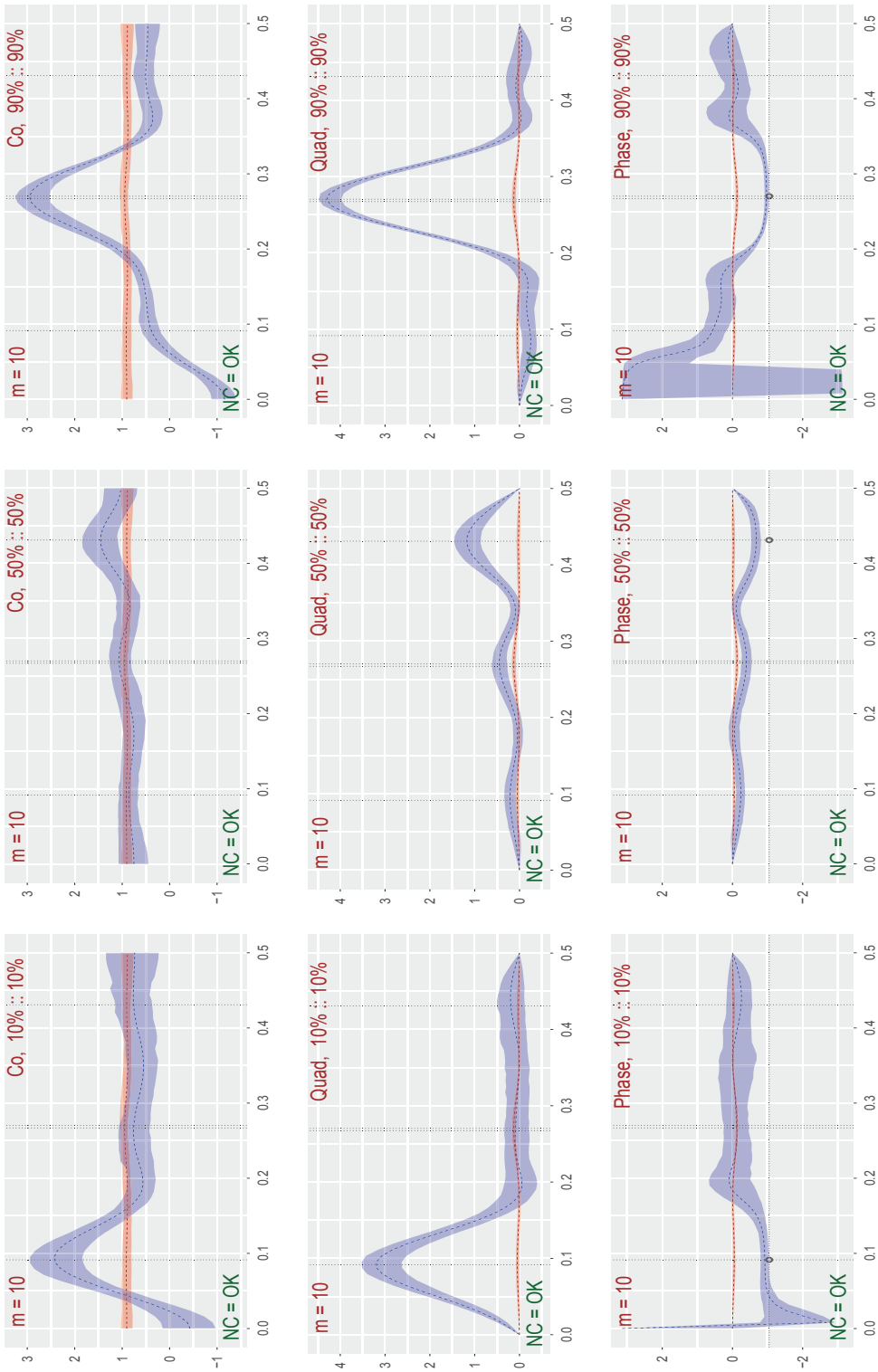

Figure 4: Plots related to eq. (3.4), common phase adjustment, designated points along the diagonal.

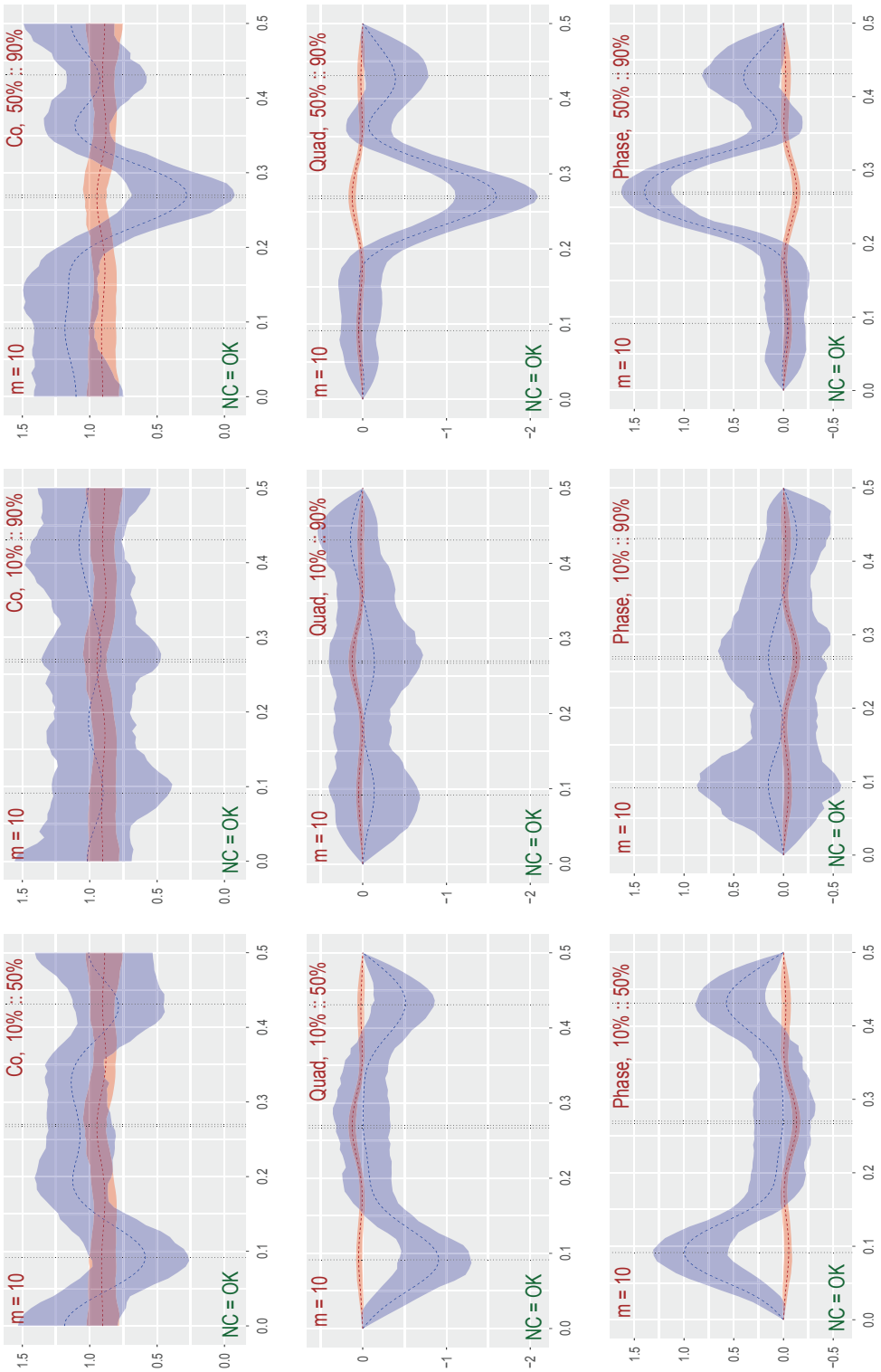

Figure 5: Plots related to eq. (3.4), common phase adjustment, behaviour away from the designated points.

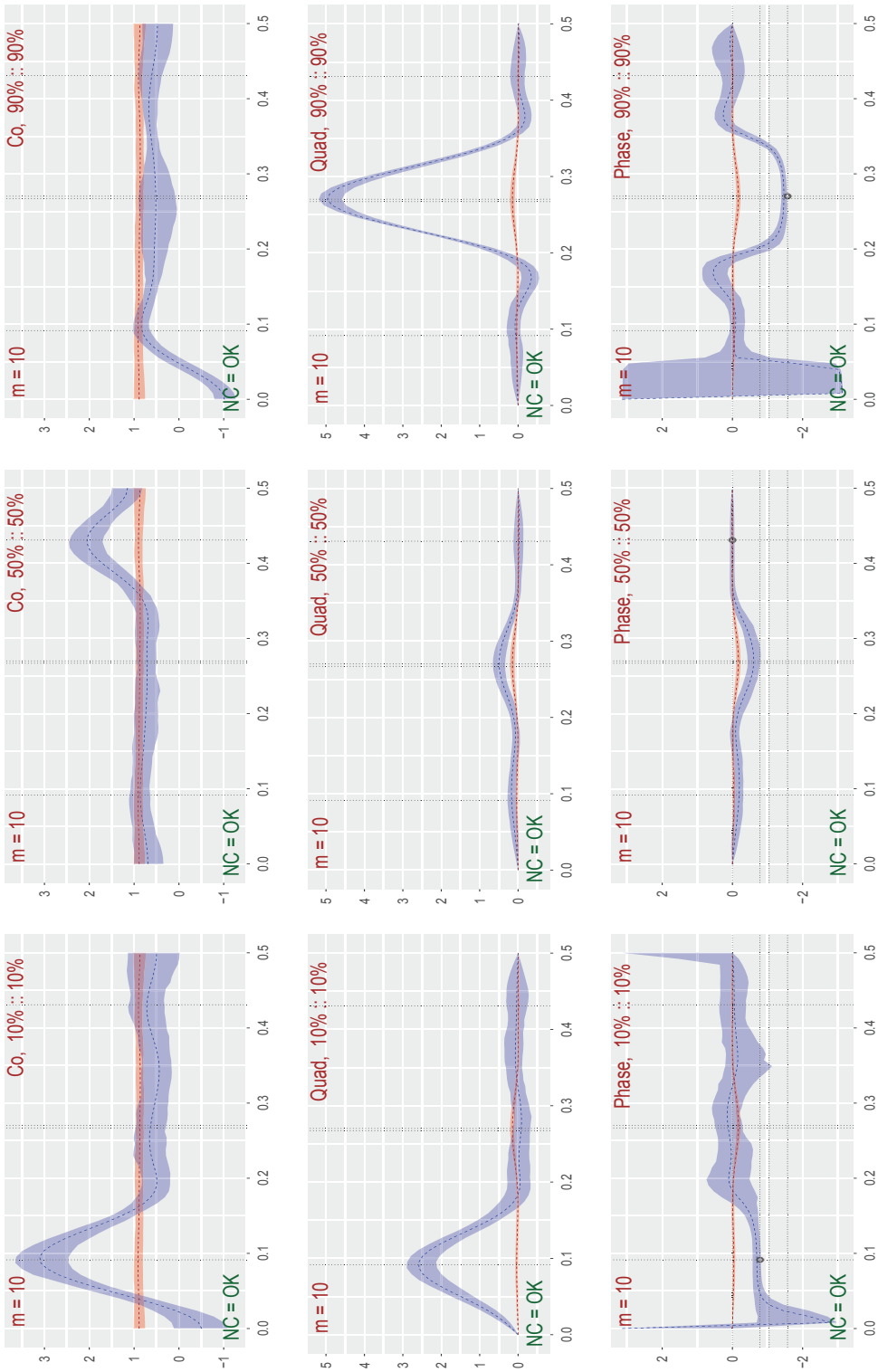

Figure 6: Plots related to eq. (3.4), individual phase adjustments, designated points along the diagonal.

3.2 A real multivariate time series and a (badly) fitted GARCH-type model

This section will show how the Co-, Quad- and Phase-plots can be used as an exploratory tool on some financial data, and then it will be seen how this approach can be used to get a visual impression of the quality of a multivariate GARCH-type model fitted to these data.

The multivariate time series sample to be considered in this section will be a bivariate subset of the (log-returns of the) tetravariate EuStockMarkets-sample from the datasets-package of R, R Core Team (2017). This data-set has been selected since it has a length that should be large enough to justify the assumption that the observed features in the Co-, Quad- and Phase-plots are not solely there due to small sample variation.

The EuStockMarkets contains 1860 daily closing prices collected in the period 1991-1998, from the following four major European stock indices: Germany DAX (Ibis), Switzerland SMI, France CAC, and UK FTSE. The data was sampled in business time, i.e., weekends and holidays was omitted.

The log-returns of the EuStockMarkets values gives a tetravariate data-set that it seems natural to model with some multivariate GARCH-type model, and the R-package rmgarch, Ghalanos (2015b), was thus used for that purpose. Note that the present paper only aims at showing how this kind of analysis can be performed, so only one very simple model was investigated – which thus gave a rather badly fitted model for the data at hand.

3.2.1 The DAX-CAC subset of the EuStockMarkets-log-returns

The log-returns of the bivariate EuStockMarkets-subset $(Y_1, Y_3) = (\text{DAX}, \text{CAC})$, of length 1859, will now be investigated. The individual pseudo-normalised traces of these observations are shown in fig. 7, and it will be from these pseudo-normalised observations that the local Gaussian cross-correlations will be computed.

The plots of the resulting m-truncated global and local spectra are shown as the red and blue lines in fig. 9, and the 90% pointwise confidence intervals has been created based on 100 block-bootstrap replicates using a block-length of 100.

No data-driven algorithm is present for the selection of the blocklength in this case, cf. section 4.1, and the rationale for the use of a given blocklength is thus based on a visual inspection of fig. 8, where the plot of the local Gaussian cross-correlations (based on the bandwidth $\boldsymbol{b} = (0.6, 0.6)$ for all the lags) are presented for the three points that will be investigated in fig. 9. It seems plausible, based on an inspection of fig. 8, that a blocklength of 100 might be needed in this case – but note that a more rigorous investigation should have been employed if this was to be an actual in depth analysis of the dependence on the blocklength.

The three points considered in fig. 9 all lie on the diagonal, since it seems easier to give an interpretation for those points. In particular, the point 10%::10% represent a situation where the market goes down both in Germany and France, whereas the points 50%::50% and 90%::90% similarly represent cases where the market either is stable or goes up in both countries.

Thus, for the purpose of the present paper, it suffices to point out that the Co-, Quad- and Phase-plots of fig. 9 indicates that the data contains nonlinear traits, which in particular is visible in the Co-plot for the point 10%::10% and for all the plots related to the point 90%::90%. It should be noted that the Co-plots at the frequency $\omega = 0$ simply gives a weighted sum of the local Gaussian cross-correlations (between $\left(Y_{1,t+h}, Y_{3,t}\right)$) seen in fig. 8, so the Co-plot peaks at

$\omega = 0$ for the points `10%::10%` and `90%::90%` are thus as expected, and the lack of a `Co`-plot peak at $\omega = 0$ for the point `50%::50%` also seems natural in view of fig. 8. It should also be noted that the $\omega = 0$ peak for the `90%::90%` `Co`-plot is lower than the corresponding peak for `10%::10%`, but this seems in this case to be due to the low truncation level used for the plots, i.e. these two peaks attain approximately the same height when a higher truncation level is applied.

It seems, for the particular parameter-configuration that generated the plots in fig. 9, to be the case that the point `90%::90%` has the most interesting `Quad`- and `Phase`-plots, but again, as noted above, it may be premature to put too much emphasis on this particular plot given the uncertainties involved in the selection of the bandwidth $\boldsymbol{b}$, the truncation level m, and the block-length to be used in the bootstrap. In particular the block-length is critical with regard to this, since a shorter block-length tends to give wider confidence-intervals, and that might result in plots where a seemingly significant difference between the local and global spectra disappears.

It should also be noted that a low number of bootstrapped replicates can be a source of small sample variation for the width of the estimated pointwise confidence intervals, and this is important to keep in mind if a minor gap is observed between the pointwise confidence intervals for the local and global spectra. Such gaps could appear or disappear when the algorithm is used to generate new computations based on the same number of bootstrapped replicates, a behaviour that in particular has been observed for the rightmost peak/trough of the `Quad`- and `Phase`-plots at the point `90%::90%` in fig. 9.

This kind of ambiguity can be countered by increasing the number of bootstrapped replicates, but that has not been done for the present example due to the uncertainty with regard to the size of the selected blocklength. In particular, if fig. 9 is recreated with the blocklength 50 instead of 100,[12] then the pointwise confidence interval for the `Phase`-plot at the point `90%::90%` widens enough to remove those occurrences where it seems that it could be a significant trough at the rightmost side. The peak at the center of the `90%::90%` `Phase`-plot does however remain significantly different from the global spectrum even with a smaller blocklength, which strengthens the impression that something of interest might be present at that frequency. However, it is important to keep in mind that this impression is based on the present combination of bandwidth and truncation level – and there are at the moment no data-driven method for the selection of these parameters. (A positive phase difference is consistent with the `90%::90%` cross-correlation plot in fig. 8, which might indicate that the DAX is leading over CAC when the market is going up.)

Remark 3.15. Note that the `shiny`-interface in the `R`-package `localgaussSpec` should be used if it is of interest to pursue a further analysis of the local Gaussian spectra of the (log-returns of the) `EuStockMarkets`-data, since that enables an interactive investigation that shows how the estimates vary based on different bandwidths $\boldsymbol{b}$ and truncation levels m, and moreover, it would also allow an investigation of how much the selection of the block-length for the bootstrap-procedure influences the widths of the pointwise confidence intervals in the `Co`-, `Quad`- and `Phase`-plots.

[12] Based on the lower panel of fig. 8, it might be ample reason to consider a block-length of 50 to be too short.

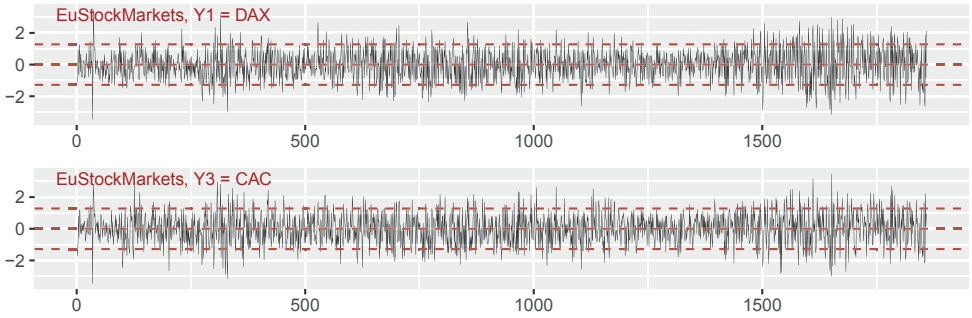

Figure 7: EuStockMarkets, pseduo-normalised components under investigation.

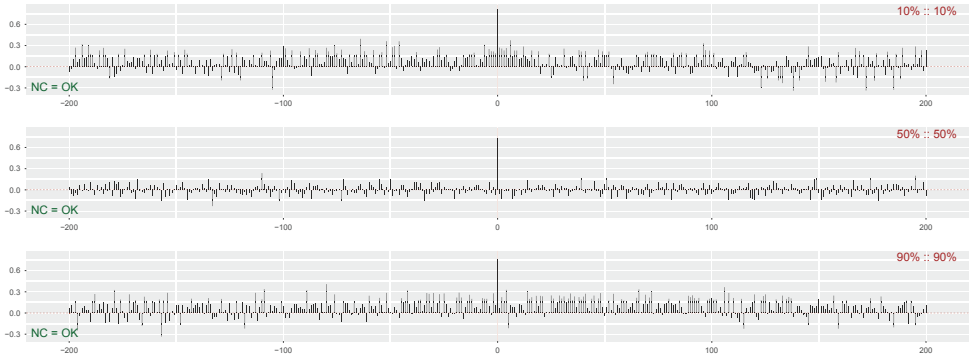

Figure 8: EuStockMarkets, local Gaussian cross-correlations between $\left(Y_{1,t+h}, Y_{3,t}\right)$, i.e. DAX is leading CAC.

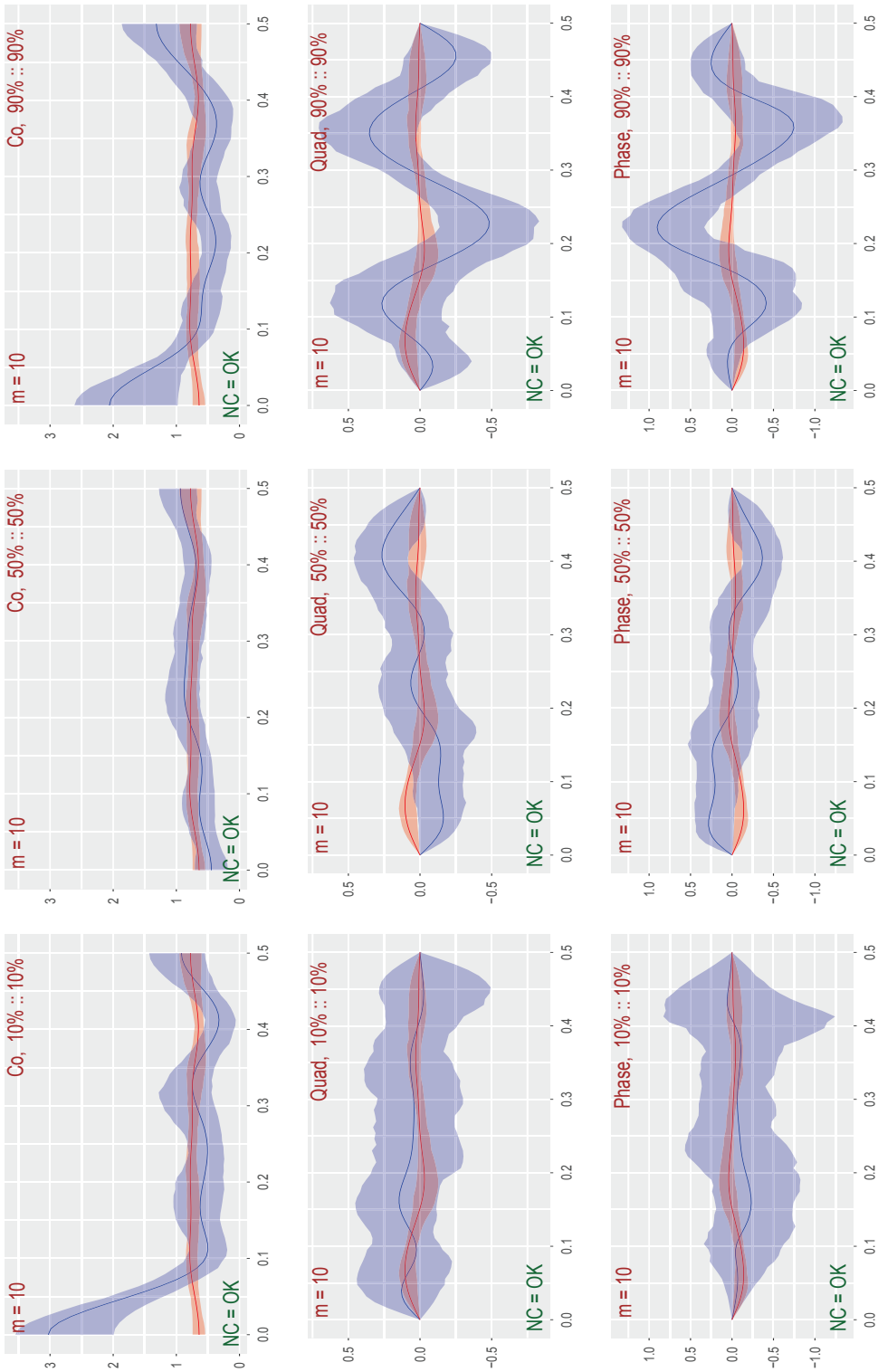

Figure 9: EuStockMarkets, local Gaussian spectra based on $\left(Y_{1,t+h}, Y_{3,t}\right)$, i.e. DAX and CAC.

3.2.2 A simple copula GARCH-model fitted to `EuStockMarkets`

It might not be obvious how to interpret the `Co-`, `Quad-` and `Phase` spectra based on the log-returns of the `EuStockMarkets`-data, but they do at least provide an approach where nonlinear dependencies might be detected from a visual inspection of the plots.

Furthermore, it is possible to use this as an exploratory tool in order to investigate whether a model fitted to the original data is capable of reproducing nonlinear traits that match those observed for the data. The procedure is straightforward:

1. Fit the selected model to the data.
2. Perform a local Gaussian spectrum investigation based on simulated samples from the fitted model. The parameters should match those used in the investigation of the original data.
3. Compare the plots based on the original data with corresponding plots based on the simulated data from the model. It can be of interest to not only compare the `Co-`, `Quad-` and `Phase`-plots, but also include plots that shows the traces and the estimated local Gaussian auto- and cross-spectra.

For the present case of interest, item 2 of the list above implies that 100 independent samples of length 1859 will be used as the basis for the construction of the `Co-`, `Quad-` and `Phase`-plots of the fitted model, and the bandwidth $\boldsymbol{b} = (0.6, 0.6)$ will be used for the estimation of the local Gaussian cross-correlations at the three points `10%::10%`, `50%::50%` and `90%::90%`.

The model: The `R`-package `rmgarch` was used to fit a simple multivariate GARCH-type model to the log-returns of the `EuStockMarkets`-data, in order to exemplify the procedure outlined above, i.e. a copula GARCH-model (cGARCH) with the simplest available univariate models for the marginals[13] was fitted to the data, and the resulting model was then used to produce figs. 10 to 12.

The traces: Figure 10 shows the pseudo-normalised trace of the Y_1- and Y_3-variables for one sample from the tetravariate cGARCH-model, and this can be compared with the corresponding pseudo-normalised trace of the DAX and CAC plot for the pseudo-normalised log-returns of the `EuStockMarkets`-data, see fig. 7 on page 27. Obviously, a comparison of one single simulated trace with the trace of the original data might not reveal much, and it should also be noted that it in general might be preferable to compare the traces before they are subjected to the pseudo-normalisation, since that could detect if the model might fail to produce sufficiently extreme outliers.

The local Gaussian correlations: Box-plots, based on the 100 independent estimates of the local Gaussian cross-correlations from the cGARCH-model, are shown in fig. 11. These can be compared with the local Gaussian cross-correlations estimated from the original sample, shown in fig. 8. It should be noted that the computational cost for the production of the box-plots in fig. 11 is substantially larger than the cost for the production of the simpler plots shown in fig. 8, so it is preferable to restrict the attention to a shorter range of lags in fig. 11. Note also that the wide range of lags included in fig. 8 is related to the need for a justification of the selected block-lenght for the bootstrap-algorithm, and it should be possible to judge the suitability of the fitted model from the shorter range of lags included in fig. 11.

[13] See Ghalanos (2015a) for details about the cGARCH-model and other options available in the `rmgarch`-package.

The impression from the lags included in fig. 11 is that the medians of the estimated local Gaussian cross-correlations for the point `50%::50%` lies quite close to zero, whereas the medians for the points `10%::10%` and `90%::90%` mostly lies slightly above zero. Almost none of the boxes for the two latter points seems to be positioned in a manner consistent with the desired outcome for a good match with the corresponding estimated values in fig. 8, and it might thus be ample reason to suspect that this cGARCH-model might better be replaced with another model instead.

The `Co-` `Quad-` and `Phase`-plots: Figure 12 shows the local Gaussian spectra for the same points $\boldsymbol{v}$ and the same configuration of parameters as those used in fig. 9. The `Co-`, `Quad-` and `Phase`-plots of fig. 12 does look like they could originate from i.i.d. white noise – which does not come as a surprise in view of the information about the local Gaussian cross-correlations in fig. 11. The `Co`-plots for the two points `10%::10%` and `90%::90%` does show that the estimates of the m-truncated local Gaussian cospectra, i.e. the blue dashed lines, might have a peak at $\omega = 0$ – but the 90% pointwise confidence intervals are too wide to support a claim that these peaks are significant. A further comparison of these two `Co`-plots with the corresponding `Co`-plots in fig. 9 (beware of different scales for the axes), does moreover show that the confidence intervals from fig. 12 are too narrow (at $\omega = 0$) to encompass the peaks observed in fig. 9 – which indicates that the selected model might be a rather bad approximation to the log-returns of the `EuStockMarkets`-data. It thus seems advisable to look for some other model instead, a natural conclusion given that no effort whatsoever was made with regard to finding reasonable marginal distributions for the copula GARCH-model used in this discussion.

Remark 3.16. It should be noted that a *local Gaussian spectra comparison* of the original data and the fitted model in practise also should include a comparison of the local Gaussian auto-spectra of the marginals, as was done in (Jordanger and Tjøstheim, 2017). These auto-spectra plots (not included in this paper) can provide some additional information useful for the model-selection process. In particular, if a model-selection algorithm for GARCH-type models has been used to pick one marginal model from a given collection of marginal models, then an investigation based on the *local Gaussian auto-spectrum* might reveal if the selected marginal model captures the local traits of the corresponding marginal observations in a reasonable manner.

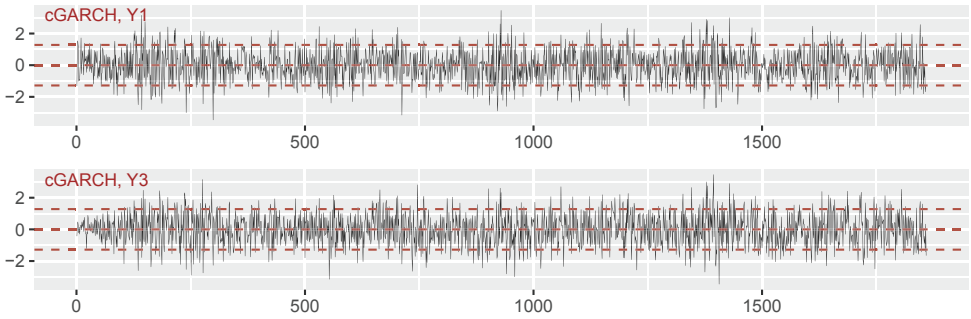

Figure 10: cGARCH, pseduo-normalised components under investigation.

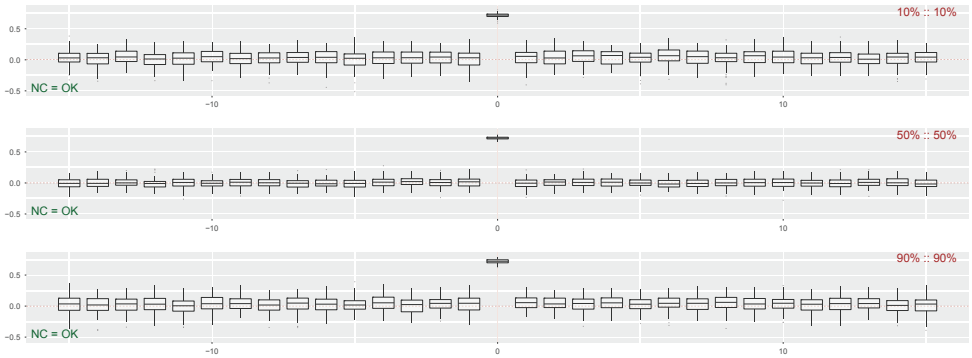

Figure 11: cGARCH, local Gaussian cross-correlations for $\left(Y_{1,t+h}, Y_{3,t}\right)$.

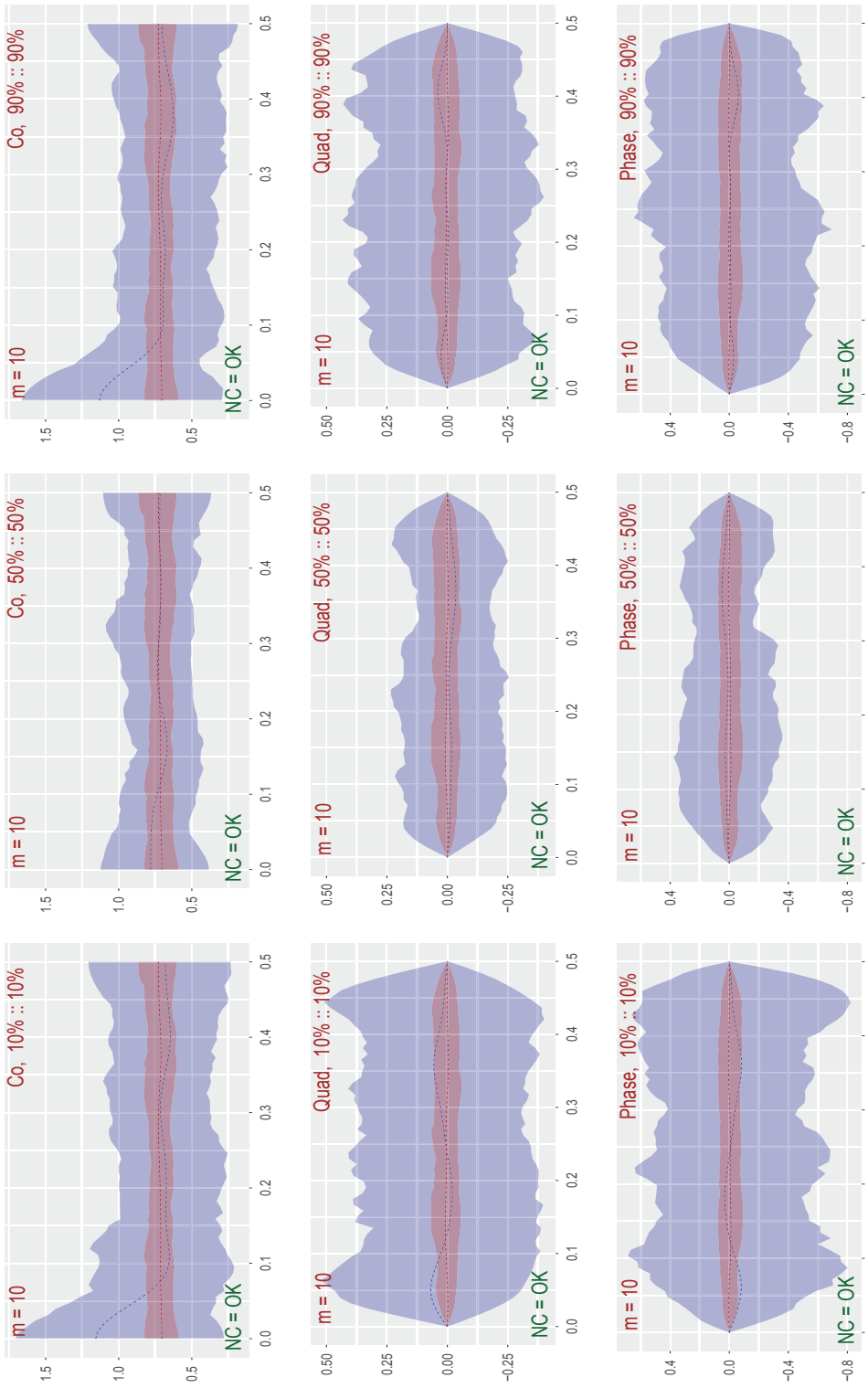

Figure 12: cGARCH, local Gaussian spectra based on $\left(Y_{1,t+h}, Y_{3,t}\right)$, corresponding to DAX and CAC of `EuStockMarkets`.

4 Discussion

This paper extends the *local Gaussian auto-spectrum* from Jordanger and Tjøstheim (2017) to the case of the *local Gaussian cross-spectrum*, and it has been seen that this can be used as an *exploratory tool* to detect non-linear traits in multivariate time series. The simulated *locally trigonometric* examples in section 3.1.2 showed that nonlinear periodicities can be detected with this machinery, but it should be noted that the selection of the point $\boldsymbol{v} = (v_1, v_2)$ was pivotal with regard to what was detected. It was also observed, see the discussion in section 3.2, that this *local Gaussian approach to spectral analysis* might be useful with regard to model selection, i.e. that it can be used to check if a model fitted to the data can reproduce local traits detected in the original sample.

This new local Gaussian approach to spectral analysis is still in its infancy, and many details have not been looked into yet. Some details, i.e. those related to the selection of the input parameters, have already been discussed in the treatment of the univariate case, see (Jordanger and Tjøstheim, 2017, section 4), and these will only be briefly discussed in this section.

4.1 The input parameters

The selection of the point $\boldsymbol{v}$ and the sample length: It is important that the coordinates v_1 and v_2 of the point $\boldsymbol{v} = (v_1, v_2)$ should not correspond to quantiles too far out in the tails, since that could allow small sample variation to dominate even when the sample itself is large,[14] cf. the discussion related to (Jordanger and Tjøstheim, 2017, fig. 2) for further details.

The selection of the bandwidth $\boldsymbol{b}$ and the truncation level m: A problem that can arise when several points $\{\boldsymbol{v}_i\}_{i=1}^r$ are investigated at the same time, is that each point $\boldsymbol{v}_i$ could have its own optimal configuration of bandwidth $\boldsymbol{b}$ and truncation level m – and the question then becomes whether these individual values should be used, or if it would be preferable to instead search for some common values for the bandwidth $\boldsymbol{b}$ and the truncation level m. The interested reader can find a more detailed discussion about this and related issues in (Jordanger and Tjøstheim, 2017, sections 4.1 and 4.2)

The block-length: It was noted in (Jordanger and Tjøstheim, 2017, section 4.3) that the standard algorithms for the selection of the bootstrap block-length depended on the (global) spectral density, and this rendered them rather useless for samples from e.g. GARCH-type models – see (Jordanger and Tjøstheim, 2017) for further details and references.

4.2 The smoothing of the estimates

The estimates of the local Gaussian auto- and cross-spectra are smoothed by the help of weighting functions $\lambda_m(h)$, that work upon the estimated local Gaussian auto- and cross-correlations, cf. definition 2.4(d). For the present investigation, the smoothing function has been inherited from the one used in the global case.

A completely different smoothing strategy can, as explained in (Jordanger and Tjøstheim, 2017, section 4.4), be based on the observation that the estimated m-truncated local Gaussian spectra corresponding to different points $\{\boldsymbol{v}_i\}_{i=1}^r$ will be jointly asymptotically normal and pairwise asymptotically independent (when $m \to \infty$ and $\boldsymbol{b} \to \boldsymbol{0}^+$ as $n \to \infty$). This implies that it should be possible to find a smoothed estimate of the local Gaussian spectral density at a given

[14] It is possible to have a point $\boldsymbol{v}$ far out in the tails if the bandwidth $\boldsymbol{b}$ is large, but the result would then hardly deserve to be referred to as an estimate of the *local* Gaussian spectral density at the point $\boldsymbol{v}$.

point $\boldsymbol{v}$, by the help of the estimated values in a grid of points surrounding $\boldsymbol{v}$. This approach resembles the one used when the ordinary global spectral densities are computed based on the periodogram,[15] but with the important distinction that the smoothing in this case would be over a grid of neighbouring points instead of a range of neighbouring frequencies.

The periodogram-approach for the estimation of the ordinary (global) auto- and cross-spectra provides an efficient estimation regime due to the *Fast Fourier Transform* (FFT), which enables the periodogram to be computed directly from the observations without the need for an explicit computation of all of the estimated auto- and cross-correlations. The effectiveness of FFT is linked to the product structure of the estimated covariance function, i.e. $\widehat{\gamma}(h) = \sum_{t=1}^{n-|h|} y_{t+|h|} \cdot y_t$, and a similar simple connection does not exist for the distribution based approach used for the local Gaussian autocorrelation. The above mentioned grid-based smoothing approach for the local Gaussian spectral density might thus be rather inefficient to use in practice, but it could still be of interest to keep this approach in mind.

4.3 The visualisation of the local Gaussian cross-spectrum

The local Gaussian cross-spectrum is complex valued, which makes a direct graphical investigation of it a bit tricky. It is possible to use a solution like the one in section 3, where a combination of the `Co-`, `Quad-` and `Phase`-plots together provide a graphical setup that enable local traits and periodicities in the data to be discovered, but other alternative might also be of interest to investigate. In particular, plots (like fig. 3) that individually might not be that informative could reveal interesting features if a sequence of them are interactively investigated by the means of a `shiny`-interface, like the one used in the `R`-package `localgaussSpec`.

A completely different approach would be to estimate the local Gaussian cross-spectra for a grid of points $\boldsymbol{v}$ in the plane, and then use three dimensional plots to see how (for a specified frequency ω) the real and imaginary parts, or the amplitude and the phase, varies over the grid. The two dimensional contour-plots corresponding to these three dimensional plots might of course also be considered, or alternative visualisations like the colour-coded phase plots from Wegert and Semmler (2011).

An obvious drawback with the above mentioned approach is that these plots might not be useful unless the number of points in the grid is (sufficiently) large, which could incur large computational costs. It might, despite this, still be of interest to initiate an investigation using such an approach with a coarse grid, since that could detect regions containing promising structures.

4.4 Possible applications

A key property of the local Gaussian spectral densities is that they for *Gaussian time series* coincide with the ordinary spectral densities. This implies that a comparison of ordinary spectra and local Gaussian spectra can detect the presence of non-Gaussian structures, which in particular can be of interest for time series whose ordinary spectra looks like white noise. Any peaks and troughs of the local Gaussian spectra can for such white noise cases naturally be considered as indicators of nonlinear structures in the time series under investigation, and it might then in particular be of interest to investigate if periodicities in the local Gaussian spectra might correspond to real phenomena in the data generating structure.

[15] See e.g. Brockwell and Davis (1986) for details.

Estimates of the local Gaussian spectral densities might also be of interest to consider as an informal tool in addition to existing tools for model selection for time series, i.e. estimates based on *repeated samples from a parametric model fitted to the original observations* can be compared with the corresponding estimates based on *the original observations*. Significant differences between these estimates (taking the pointwise confidence intervals into account) could indicate that the selected model might not be good enough, and that it thus might be prudent to look for a better model.

These applications of the local Gaussian spectra could provide useful insight even when it is unknown whether or not the investigated time series actually satisfies the requirements needed for the asymptotic theory to work. It might however occur problems due to the (at the present state of development) unresolved issues regarding the selection of the input parameters – in particular, cases can be encountered where it is hard to figure out if a peak/trough represents an actual nonlinear phenomenon, or if it is small-sample variation due to an unsavoury combination of input parameters. More work is needed in order to resolve these issues.

5 Conclusion

The *local Gaussian cross-spectrum* $f_{k\ell;\mathfrak{v}|p}(\omega)$ can be used to detect nonlinear dependencies between the marginals $Y_{k,t}$ and $Y_{\ell,t}$ of a multivariate time series $\left\{\boldsymbol{Y}_t = \left(Y_{1,t}, \ldots, Y_{d,t}\right)\right\}_{t\in\mathbb{Z}}$. This provides, together with the *local Gaussian auto-spectrum* from Jordanger and Tjøstheim (2017), a *local Gaussian approach to spectral analysis*, which in particular can be used to detect if the time series under investigation deviates from being Gaussian. For time series whose ordinary auto- and cross-spectra are flat, any peaks and troughs from the *local Gaussian approach* can then be considered indicators of *local nonlinear traits* and *local periodicities*.

Appendix A: The proofs of theorems 2.10 to 2.12

This appendix presents the proofs of the asymptotic results stated in the main part of the paper. The proof of the result for the m-truncated estimate of the local Gaussian cross-spectrum $f_{k\ell:\boldsymbol{v}|p}(\omega)$ is in essence identical to the one encountered in Jordanger and Tjøstheim (2017) for the local Gaussian auto-spectrum $f_{kk:\boldsymbol{v}|p}(\omega)$, whereas the proofs for the estimates of the local Gaussian amplitude- and phase-spectra are identical in structure to those encountered in the ordinary global case. Some technical details needed for the proof of theorem 2.10 are covered in appendix B.

Proof of theorem 2.10, page 11.
The case $\omega \notin \frac{1}{2}\cdot\mathbb{Z}$ will be treated first, since the other case follows from a trivial adjustment of the setup. The key observation for this case is that the sum that defines $\widehat{f}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega)$, see eq. (2.6) in definition 2.4(d), implies that $\widehat{c}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega)$ and $\widehat{q}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega)$ can be realised as the following inner products,

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

$$\widehat{q}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega) = \boldsymbol{\Lambda}'_{q|\overline{\boldsymbol{m}}}(\omega)\cdot\widehat{\boldsymbol{P}}_{k\ell:\overline{\boldsymbol{m}}|\boldsymbol{b}|p}(\boldsymbol{v},\breve{\boldsymbol{v}})\,, \tag{A.1b}$$

where $\boldsymbol{\Lambda}'_{c|\overline{\boldsymbol{m}}}(\omega)$ and $\boldsymbol{\Lambda}'_{q|\overline{\boldsymbol{m}}}(\omega)$ respectively contains the coefficients $\lambda_{m}(h)\cos(2\pi\omega h)$ and $\lambda_{m}(h)\sin(2\pi\omega h)$, and where $\widehat{\boldsymbol{P}}_{k\ell:\overline{\boldsymbol{m}}|\boldsymbol{b}|p}(\boldsymbol{v},\breve{\boldsymbol{v}}) = \big[\widehat{\rho}_{\ell k:\breve{\boldsymbol{v}}|p}(m),\ldots,\widehat{\rho}_{\ell k:\breve{\boldsymbol{v}}|p}(1),\widehat{\rho}_{k\ell:\boldsymbol{v}|p}(0),\ldots,\widehat{\rho}_{k\ell:\boldsymbol{v}|p}(m)\big]'$ contains the $2m+1$ estimates of the local Gaussian cross-correlations (based on the bandwidth $\boldsymbol{b}$). The vector $\widehat{\boldsymbol{P}}_{k\ell:\overline{\boldsymbol{m}}|\boldsymbol{b}|p}(\boldsymbol{v},\breve{\boldsymbol{v}})$ can by the help of a suitable $(2m+1)\times(2m+1)p$ matrix $\boldsymbol{E}'_{\overline{\boldsymbol{m}}|p}$ (based on the vectors $\boldsymbol{e}'_{p}$ that gives $\rho_{k\ell:\boldsymbol{v}|p}(h) = \boldsymbol{e}'_{p}\cdot\boldsymbol{\theta}_{\boldsymbol{v}|k\ell:h|p}$) be expressed as

$$\widehat{\boldsymbol{P}}_{k\ell:\overline{\boldsymbol{m}}|\boldsymbol{b}|p}(\boldsymbol{v},\breve{\boldsymbol{v}}) = \boldsymbol{E}'_{\overline{\boldsymbol{m}}|p}\cdot\widehat{\boldsymbol{\Theta}}_{k\ell:\overline{\boldsymbol{m}}|\boldsymbol{b}|p}(\boldsymbol{v},\breve{\boldsymbol{v}})\,, \tag{A.2}$$

where $\widehat{\boldsymbol{\Theta}}_{k\ell:\overline{\boldsymbol{m}}|\boldsymbol{b}|p}(\boldsymbol{v},\breve{\boldsymbol{v}}) = \Big[\widehat{\boldsymbol{\theta}}_{\breve{\boldsymbol{v}}|\ell k:m:n|p},\ldots,\widehat{\boldsymbol{\theta}}_{\breve{\boldsymbol{v}}|\ell k:1:n|p},\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|k\ell:0:n|p},\ldots,\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|k\ell:m:n|p}\Big]'$ is the length $(2m+1)p$ vector created by stacking into one vector all the estimated parameters from the local Gaussian approximations. It follows from this that the target of interest can be written as

$$\begin{bmatrix}\widehat{c}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega)\\ \widehat{q}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega)\end{bmatrix} = \begin{bmatrix}\boldsymbol{\Lambda}'_{c|\overline{\boldsymbol{m}}}(\omega)\\ \boldsymbol{\Lambda}'_{q|\overline{\boldsymbol{m}}}(\omega)\end{bmatrix}\cdot\boldsymbol{E}'_{\overline{\boldsymbol{m}}|p}\cdot\widehat{\boldsymbol{\Theta}}_{k\ell:\overline{\boldsymbol{m}}|\boldsymbol{b}|p}(\boldsymbol{v},\breve{\boldsymbol{v}})\,, \tag{A.3}$$

which together with the asymptotic normality result from theorem B.3 (page 39), i.e.

$$\sqrt{n(b_1b_2)^{(p+1)/2}}\cdot\Big(\widehat{\boldsymbol{\Theta}}_{k\ell:\overline{\boldsymbol{m}}|\boldsymbol{b}|p}(\boldsymbol{v},\breve{\boldsymbol{v}}) - \boldsymbol{\Theta}_{k\ell:\overline{\boldsymbol{m}}|\boldsymbol{b}|p}(\boldsymbol{v},\breve{\boldsymbol{v}})\Big) \xrightarrow{\;d\;} \mathrm{N}\big(\boldsymbol{0},\Sigma_{\boldsymbol{v}|k\ell:\overline{\boldsymbol{m}}|p}\big)\,, \tag{A.4}$$

and Brockwell and Davis (1986, proposition 6.4.2, p. 211) gives that

$$\sqrt{n(b_1b_2)^{(p+1)/2}/m}\cdot\left(\begin{bmatrix}\widehat{c}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega)\\ \widehat{q}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega)\end{bmatrix} - \begin{bmatrix}c_{k\ell:\boldsymbol{v}|p}(\omega)\\ q_{k\ell:\boldsymbol{v}|p}(\omega)\end{bmatrix}\right) \tag{A.5}$$

is asymptotically bivariate normally distributed with mean $\mathbf{0}$ and covariance matrix

$$\frac{1}{m} \cdot \left(\begin{bmatrix} \boldsymbol{\Lambda}'_{c|\overline{\underline{m}}}(\omega) \\ \boldsymbol{\Lambda}'_{q|\overline{\underline{m}}}(\omega) \end{bmatrix} \cdot \boldsymbol{E}'_{\overline{\underline{m}}|p} \cdot \Sigma_{\boldsymbol{v}|k\ell:\overline{\underline{m}}|p} \cdot \boldsymbol{E}_{\overline{\underline{m}}|p} \cdot \left[\boldsymbol{\Lambda}_{c|\overline{\underline{m}}}, \boldsymbol{\Lambda}_{q|\overline{\underline{m}}} \right] \right). \tag{A.6}$$

The specified form of the covariance matrix given in theorem 2.10 now follows by the help of some linear algebra, the observation in theorem B.3 that

$$\Sigma_{\boldsymbol{v}|k\ell:\overline{\underline{m}}|p} \coloneqq \left(\bigoplus_{h=m}^{1} \Sigma_{\breve{\boldsymbol{v}}|\ell k:h|p} \right) \bigoplus \left(\bigoplus_{h=0}^{m} \Sigma_{\boldsymbol{v}|k\ell:h|p} \right), \tag{A.7}$$

and the definition $\tilde{\sigma}^2_{\boldsymbol{v}|k\ell|p}(h) \coloneqq \boldsymbol{e}'_p \cdot \Sigma_{\boldsymbol{v}|k\ell:h|p} \cdot \boldsymbol{e}_p$ from theorem B.1.

It is easy to see that both $\sigma^2_{c|k\ell:\boldsymbol{v}|k\ell|p}(\omega)$ and $\sigma^2_{q|k\ell:\boldsymbol{v}|k\ell|p}(\omega)$ from eq. (2.19) are nonzero when $\omega \notin \frac{1}{2} \cdot \mathbb{Z}$, as required for the validity of (Brockwell and Davis, 1986, proposition 6.4.2, p. 211). The proof for the case $\omega \in \frac{1}{2} \cdot \mathbb{Z}$ can be constructed in the same manner, simply ignoring the components having sine-terms. □

The key observation for the proof of theorems 2.11 and 2.12 is that they both follow as a consequence of theorem 2.10 and Brockwell and Davis (1986, proposition 6.4.3, p. 211). Note that these arguments are quite similar to those used for the investigation of the estimates of the ordinary amplitude and phase spectra in (Brockwell and Davis, 1986, p.448-449).

Proof of theorem 2.11, page 12.
First observe that the function $h(x_1, x_2) = \sqrt{x_1^2 + x_2^2}$ implies that

$$\widehat{\alpha}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega) - \alpha_{k\ell:\boldsymbol{v}|p}(\omega) = h\big(\widehat{c}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega), \widehat{q}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega)\big) - h\big(c_{k\ell:\boldsymbol{v}|p}(\omega), q_{k\ell:\boldsymbol{v}|p}(\omega)\big), \tag{A.8}$$

and then observe that the asymptotic covariance matrix in theorem 2.10

$$\Sigma_{k\ell:\boldsymbol{v}|p}(\omega) \coloneqq \begin{pmatrix} \sigma^2_{c|k\ell:\boldsymbol{v}|p}(\omega) & 0 \\ 0 & \sigma^2_{q|k\ell:\boldsymbol{v}|p}(\omega) \end{pmatrix}, \tag{A.9}$$

obviously is a symmetric non-negative definite matrix.

It now follows from Brockwell and Davis (1986, proposition 6.4.2, p. 211) that

$$\sqrt{n(b_1 b_2)^{(p+1)/2}/m} \cdot \left\{ h\big(\widehat{c}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega), \widehat{q}^{\,m}_{k\ell:\boldsymbol{v}|p}(\omega)\big) - h\big(c_{k\ell:\boldsymbol{v}|p}(\omega), q_{k\ell:\boldsymbol{v}|p}(\omega)\big) \right\} \xrightarrow{\;d\;} \mathrm{N}(0, \sigma^2_\alpha(\omega)), \tag{A.10}$$

where $\sigma^2_\alpha(\omega) = \boldsymbol{D} \cdot \Sigma_{k\ell:\boldsymbol{v}|p}(\omega) \cdot \boldsymbol{D}'$, with

$$\boldsymbol{D} = \left[\frac{\partial}{\partial x_1} h(x_1, x_2), \frac{\partial}{\partial x_2} h(x_1, x_2) \right] = \left[x_1/\sqrt{x_1^2 + x_2^2}, x_2/\sqrt{x_1^2 + x_2^2} \right] \tag{A.11}$$

evaluated in $(x_1, x_2) = \big(c_{k\ell:\boldsymbol{v}|p}(\omega), q_{k\ell:\boldsymbol{v}|p}(\omega)\big)$.

A simple calculation gives $\boldsymbol{D} = \big[c_{k\ell:\boldsymbol{v}|p}(\omega)/\alpha_{k\ell:\boldsymbol{v}|p}(\omega), q_{k\ell:\boldsymbol{v}|p}(\omega)/\alpha_{k\ell:\boldsymbol{v}|p}(\omega)\big]$, from which it follows that $\sigma^2_\alpha(\omega) = \big(c^2_{k\ell:\boldsymbol{v}|p}(\omega) \cdot \sigma^2_{c|k\ell:\boldsymbol{v}|p}(\omega) + q^2_{k\ell:\boldsymbol{v}|p}(\omega) \cdot \sigma^2_{q|k\ell:\boldsymbol{v}|p}(\omega)\big) / \alpha^2_{k\ell:\boldsymbol{v}|p}(\omega)$. □

Proof of theorem 2.12, page 12.
This argument is quite similar to the proof of theorem 2.11, and only the details that are different will thus be included. In this case the function of interest is $h(x_1, x_2) = \tan^{-1}(x_2/x_1)$, from which it follows that

$$\left[\frac{\partial}{\partial x_1}h(x_1,x_2), \frac{\partial}{\partial x_2}h(x_1,x_2)\right] = \left[-x_2/\left(x_1^2+x_2^2\right), x_1/\left(x_1^2+x_2^2\right)\right]. \tag{A.12}$$

This implies that $\boldsymbol{D} = \left[-q_{k\ell:\boldsymbol{v}|p}(\omega)/\alpha^2_{k\ell:\boldsymbol{v}|p}(\omega), c_{k\ell:\boldsymbol{v}|p}(\omega)/\alpha^2_{k\ell:\boldsymbol{v}|p}(\omega)\right]$ and a simple calculation now gives

$$\sigma^2_\phi(\omega) = \left(q^2_{k\ell:\boldsymbol{v}|p}(\omega)\cdot\sigma^2_{c|k\ell:\boldsymbol{v}|p}(\omega) + c^2_{k\ell:\boldsymbol{v}|p}(\omega)\cdot\sigma^2_{q|k\ell:\boldsymbol{v}|p}(\omega)\right)/\alpha^4_{k\ell:\boldsymbol{v}|p}(\omega), \tag{A.13}$$

which completes the proof. □

Appendix B: The underlying asymptotic results

B.1 The bivariate case, a brief overview and the $\widehat{\rho}_{k\ell:\boldsymbol{v}|p}(h)$-case

The main ingredient for the theoretical setup is a translation of the bivariate results from (Tjøstheim and Hufthammer, 2013) into the multivariate framework, and this is almost identical to the discussion that was given in (Jordanger and Tjøstheim, 2017, appendix B.1.2). The main difference is that two extra indices (k and ℓ) now are needed in order to specify which components from $\boldsymbol{Y}_t = \left(Y_{1,t}, \ldots, Y_{d,t}\right)$ that are investigated.

The basic building-blocks was given in section 2.5.1, see definitions 2.6 to 2.9, and the first target of interest is to define a suitable bivariate penalty function relative to the requirement of eq. (2.10). For a sample of size n from $\{\boldsymbol{Y}_{k\ell:h:t}\}_{t\in\mathbb{Z}}$, and with the present notation, the local penalty function from (Tjøstheim and Hufthammer, 2013) can be described as

$$\begin{aligned} Q_{\boldsymbol{v}|k\ell:h:n|p}\left(\boldsymbol{\theta}_{\boldsymbol{v}|k\ell:h|p}\right) := &-\sum_{t=1}^{n} K_{k\ell:h:\boldsymbol{b}}(\boldsymbol{Y}_{k\ell:h:t}-\boldsymbol{v})\log\psi_p\left(\boldsymbol{Y}_{k\ell:h:t};\boldsymbol{\theta}_{\boldsymbol{v}|k\ell:h|p}\right) \\ &+ n\int_{\mathbb{R}^2} K_{k\ell:h:\boldsymbol{b}}(\boldsymbol{y}_{k\ell:h}-\boldsymbol{v})\,\psi_p\left(\boldsymbol{y}_{k\ell:h};\boldsymbol{\theta}_{\boldsymbol{v}|k\ell:h|p}\right)\mathrm{d}\boldsymbol{y}_{k\ell:h}, \end{aligned} \tag{B.1}$$

and from this, under suitable regularity conditions and by the help of the Klimko-Nelson approach, the following asymptotic normality result can be obtained for the estimated parameters,

$$\sqrt{n(b_1b_2)^{(p+1)/2}}\cdot\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|k\ell:h:n|p}-\boldsymbol{\theta}_{\boldsymbol{v}|k\ell:h|p}\right)\xrightarrow{d}\mathrm{N}\left(\boldsymbol{0},\Sigma_{\boldsymbol{v}|k\ell:h|p}\right). \tag{B.2}$$

See Tjøstheim and Hufthammer (2013, Th. 3) for the details for the $p=5$ case, and note that the $p=1$ case follows from a simplification of that result.

Remark B.1. For the $p=1$ case, $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|k\ell:h:n|1}$ consists of the single value $\widehat{\rho}_{k\ell:\boldsymbol{v}|1}(h)$, and eq. (B.2) gives the univariate result of interest. For the $p=5$ case, $\widehat{\rho}_{k\ell:\boldsymbol{v}|5}(h)$ can be expressed as $\boldsymbol{e}_5'\cdot\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|k\ell:h:n|5}$, where $\boldsymbol{e}_5'$ is the unit vector that picks out the correlation from $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|k\ell:h:n|5}$, and it follows

from (Brockwell and Davis, 1986, proposition 6.4.2, p. 211) that

$$\sqrt{n(b_1 b_2)^3} \cdot \big(\widehat{\rho}_{k\ell:\boldsymbol{v}|5}(h) - \rho_{k\ell:\boldsymbol{v}|5}(h)\big) \xrightarrow{d} \mathrm{N}\big(0, \boldsymbol{e}_5' \cdot \Sigma_{\boldsymbol{v}|k\ell:h|5} \cdot \boldsymbol{e}_5\big)\,. \tag{B.3}$$

These observations can (by introducing $\boldsymbol{e}_1 \coloneqq 1$) be collected in the following result, which is included in order to give a reference for the statements in theorem 2.10.

Theorem B.1. *Under assumptions 2.1 to 2.3, the following univariate asymptotic result holds for the estimates $\widehat{\rho}_{k\ell:\boldsymbol{v}|p}(h)$ of the local Gaussian cross-correlations $\rho_{k\ell:\boldsymbol{v}|p}(h)$.*

$$\sqrt{n(b_1 b_2)^{(p+1)/2}} \cdot \big(\widehat{\rho}_{k\ell:\boldsymbol{v}|p}(h) - \rho_{k\ell:\boldsymbol{v}|p}(h)\big) \xrightarrow{d} \mathrm{N}\big(0, \tilde{\sigma}^2_{\boldsymbol{v}|k\ell|p}(h)\big)\,, \tag{B.4}$$

where $\tilde{\sigma}^2_{\boldsymbol{v}|k\ell|p}(h) \coloneqq \boldsymbol{e}_p' \cdot \Sigma_{\boldsymbol{v}|k\ell:h|p} \cdot \boldsymbol{e}_p$.

B.2 The asymptotic result for $\widehat{\Theta}_{k\ell:\overline{\boldsymbol{m}}|\boldsymbol{b}|p}(\boldsymbol{v}, \breve{\boldsymbol{v}})$

The Klimko-Nelson approach from the bivariate case can be extended to the present case of interest in the same manner as it was done for the local Gaussian auto-spectrum in (Jordanger and Tjøstheim, 2017, appendix B).

Definition B.2. *For each bivariate penalty function $Q_{\boldsymbol{v}|k\ell:h:n|p}\big(\boldsymbol{\theta}_{\boldsymbol{v}|k\ell:h|p}\big)$ (as given in eq. (B.1)), denote by $\widetilde{Q}_{\boldsymbol{v}|k\ell:h:n|p}\big(\boldsymbol{\theta}_{\boldsymbol{v}|k\ell:h|p}\big)$ the extension of it from a function of $\boldsymbol{Y}_{k\ell:h:t} \coloneqq \big[Y_{k,t+h}, Y_{\ell,t}\big]'$ to a function of $\big[Y_{k,t+m}, \ldots, Y_{k,t}, Y_{\ell,t+m}, \ldots, Y_{\ell,t}\big]'$. Use these extensions do define the new penalty function*

$$Q_{\boldsymbol{v}:\breve{\boldsymbol{v}}|k\ell:\overline{\boldsymbol{m}}:n|p}\big(\boldsymbol{\Theta}_{k\ell:\overline{\boldsymbol{m}}|\boldsymbol{b}|p}(\boldsymbol{v}, \breve{\boldsymbol{v}})\big) \coloneqq \sum_{h=m}^{1} \widetilde{Q}_{\breve{\boldsymbol{v}}|\ell k:h:n|p}\big(\boldsymbol{\theta}_{\breve{\boldsymbol{v}}|\ell k:h|p}\big) + \sum_{h=0}^{m} \widetilde{Q}_{\boldsymbol{v}|k\ell:h:n|p}\big(\boldsymbol{\theta}_{\boldsymbol{v}|k\ell:h|p}\big)\,. \tag{B.5}$$

The $2m+1$ bivariate components in the sum that defines $Q_{\boldsymbol{v}:\breve{\boldsymbol{v}}|k\ell:\overline{\boldsymbol{m}}:n|p}\Big(\widehat{\boldsymbol{\Theta}}_{k\ell:\overline{\boldsymbol{m}}|\boldsymbol{b}|p}(\boldsymbol{v}, \breve{\boldsymbol{v}})\Big)$ have no common parameters, so the optimisation of the parameters for the different summands can be performed independently. The optimal parameter vector $\widehat{\boldsymbol{\Theta}}_{k\ell:\overline{\boldsymbol{m}}|\boldsymbol{b}|p}(\boldsymbol{v}, \breve{\boldsymbol{v}})$ for the penalty function $Q_{\boldsymbol{v}:\breve{\boldsymbol{v}}|k\ell:\overline{\boldsymbol{m}}:n|p}$ (for a given sample) can thus be constructed by stacking on top of each other the parameter vectors $\widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|k\ell:h|p}$ and $\widehat{\boldsymbol{\theta}}_{\breve{\boldsymbol{v}}|\ell k:h|p}$ that optimise the individual summands in eq. (B.5).

The Klimko-Nelson approach can now be used on the penalty function from eq. (B.5) i.e. four requirements related to the penalty function must be verified before the desired asymptotic result for the paramater-vector $\widehat{\boldsymbol{\Theta}}_{k\ell:\overline{\boldsymbol{m}}|\boldsymbol{b}|p}(\boldsymbol{v}, \breve{\boldsymbol{v}})$ is obtained. The following cross-spectrum analogue of (Jordanger and Tjøstheim, 2017, theorem B.23) can now be stated for the present case of interest.

Theorem B.3. *Under assumptions 2.1 to 2.3, the following asymptotic behaviour holds for the estimated parameters $\widehat{\boldsymbol{\Theta}}_{k\ell:\overline{\boldsymbol{m}}|\boldsymbol{b}|p}(\boldsymbol{v}, \breve{\boldsymbol{v}}) = \Big[\widehat{\boldsymbol{\theta}}_{\breve{\boldsymbol{v}}|\ell k:m:n|p}, \ldots, \widehat{\boldsymbol{\theta}}_{\breve{\boldsymbol{v}}|\ell k:1:n|p}, \widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|k\ell:0:n|p}, \ldots, \widehat{\boldsymbol{\theta}}_{\boldsymbol{v}|k\ell:m:n|p}\Big]'$,*

$$\sqrt{n(b_1 b_2)^{(p+1)/2}} \cdot \Big(\widehat{\boldsymbol{\Theta}}_{k\ell:\overline{\boldsymbol{m}}|\boldsymbol{b}|p}(\boldsymbol{v}, \breve{\boldsymbol{v}}) - \boldsymbol{\Theta}_{k\ell:\overline{\boldsymbol{m}}|\boldsymbol{b}|p}(\boldsymbol{v}, \breve{\boldsymbol{v}})\Big) \xrightarrow{d} \mathrm{N}\big(\boldsymbol{0}, \Sigma_{\boldsymbol{v}|k\ell:\overline{\boldsymbol{m}}|p}\big)\,, \tag{B.6}$$

where the matrix $\Sigma_{\boldsymbol{v}|k\ell:\overline{m}|p}$ *is the direct sum of the matrices from eq.* (B.2) *that occurs when the individual bivariate components of the penalty function is investigated, i.e.*

$$\Sigma_{\boldsymbol{v}|k\ell:\overline{m}|p} \coloneqq \left(\bigoplus_{h=m}^{1} \Sigma_{\breve{\boldsymbol{v}}|\ell k:h|p}\right) \bigoplus \left(\bigoplus_{h=0}^{m} \Sigma_{\boldsymbol{v}|k\ell:h|p}\right). \tag{B.7}$$

Proof. This result follows when the Klimko-Nelson approach is used with the local penalty-function $Q_{\boldsymbol{v}:\breve{\boldsymbol{v}}|k\ell:\overline{m}:n|p}\big(\boldsymbol{\Theta}_{k\ell:\overline{m}|\boldsymbol{b}|p}(\boldsymbol{v},\breve{\boldsymbol{v}})\big)$ from eq. (B.5), and the proof is in essence identical to the proof of (Jordanger and Tjøstheim, 2017, theorem B.23). The three first requirements of the Klimko-Nelson approach follows trivially from the corresponding investigation for the bivariate case, whereas the proof of the fourth requirement must take into account how $m \to \infty$ and $\boldsymbol{b} \to \mathbf{0}^{+}$ as $n \to \infty$.

The investigation of the fourth requirement of the Klimko-Nelson approach can be done in the exact same manner that was employed in Jordanger and Tjøstheim (2017), i.e. first construct a collection of simple random variables whose interaction and asymptotic properties are easy to investigate, then use these basic building blocks to construct a more complicated random variable $\mathfrak{Q}^{n}_{\boldsymbol{v}|\overline{m}}$ that has the same limiting distribution as the estimator of $\sqrt{b_1 b_2}\boldsymbol{\nabla}_{k\ell:\overline{m}|p} Q_{\boldsymbol{v}:\breve{\boldsymbol{v}}|k\ell:\overline{m}:n|p}\big(\boldsymbol{\Theta}_{k\ell:\overline{m}|\boldsymbol{b}|p}(\boldsymbol{v},\breve{\boldsymbol{v}})\big)$ (where $\boldsymbol{\nabla}_{k\ell:\overline{m}|p}$ is obtained by stacking together $\boldsymbol{\nabla}_{k\ell:h|p}$). After this, it is sufficient to use standard methods to prove that the limiting distribution of $\mathfrak{Q}^{n}_{\boldsymbol{v}|\overline{m}}$ is the desired multivariate normal distribution, and the statement for the parameter vectors then follows from the Klimko-Nelson theorem and some linear algebra. □

References

Berentsen, G. D., Kleppe, T. S., Tjøstheim, D. B., Feb. 2014. 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

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.

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

Chung, J., Hong, Y., 2007. Model-free evaluation of directional predictability in foreign exchange markets. Journal of Applied Econometrics 22 (5), 855–889.
URL http://dx.doi.org/10.1002/jae.965

Ghalanos, A., 2015a. The rmgarch models: Background and properties. (Version 1.3-0).
URL https://CRAN.R-project.org/web/packages/rmgarch/vignettes/The_rmgarch_models.pdf

Ghalanos, A., 2015b. rmgarch: Multivariate GARCH models. R package version 1.3-0.
URL https://CRAN.R-project.org/package=rmgarch

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., 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

Jordanger, L. A., Tjøstheim, D., 2017. Nonlinear spectral analsys via the local Gaussian correlation.
URL http://arxiv.org/abs/1708.02166v1

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

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

R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL https://www.R-project.org/

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

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.

Wegert, E., Semmler, G., June/July 2011. Phase Plots of Complex Functions: A Journey in Illustration. Notices of the AMS 58 (6), 768–780.
URL http://www.ams.org/journals/notices/201106/rtx110600768p.pdf

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/

Paper III

2.3 Model selection of copulas: AIC versus a cross validation copula information criterion

Lars Arne Jordanger and Dag Tjøstheim

Statistics and Probability Letters, **92**, 49-255 (2014)

Statistics and Probability Letters 92 (2014) 249–255

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Model selection of copulas: AIC versus a cross validation copula information criterion

Lars Arne Jordanger*, Dag Tjøstheim

University of Bergen, Department of Mathematics, P.O. Box 7800, N-5020 Bergen, Norway

ARTICLE INFO

Article history:
Received 29 January 2014
Received in revised form 4 June 2014
Accepted 5 June 2014
Available online 12 June 2014

Keywords:
Model selection
AIC
Copula information criterion
CIC
Pseudo-observations

ABSTRACT

Akaike Information Criterion (AIC) is frequently employed in the semiparametric setting of selection of copula models, even though as a model selection tool it was developed in a parametric setting. Recently a Copula Information Criterion (CIC) has been especially designed for copula model selection. In this paper we examine the two approaches and present a simulation study where the performance of a cross-validated version of CIC is compared with the AIC criterion. Only minor differences are observed.

1. Introduction

For a set of *independent* observations $\mathcal{X}_n = \{\boldsymbol{x}_i\}_{i=1}^n$, $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{id})$, and a proposed parametric family F with parameters $\boldsymbol{\alpha}$, the Akaike Information Criterion, introduced in Akaike (1974), will assign the following value to the members of the family F,[1]

$$\mathrm{AIC}_F = 2\ell_F(\widehat{\boldsymbol{\alpha}}) - 2\dim(\boldsymbol{\alpha}), \tag{1}$$

where ℓ_F is the log-likelihood-function, $\widehat{\boldsymbol{\alpha}}$ is the maximum-likelihood estimate of $\boldsymbol{\alpha}$ based on the independent observations $\mathcal{X}_n$, and $\dim(\boldsymbol{\alpha})$ is the number of parameters in the model.

AIC was introduced in the setting of time-series, but it is also formally valid in more general situations. It is customary to apply a modified version of AIC when the goal is to find the (parametric) copula model C that best describes the set of pseudo-observations ${}^p\mathcal{X}_n$ in the unit (hyper)cube, obtained when the empirical marginal distributions constructed from the observations $\mathcal{X}_n$ are used. The modified version of AIC on ${}^p\mathcal{X}_n$ will in this article be denoted by ${}^p\mathrm{AIC}$,[2] and it is given by

$${}^p\mathrm{AIC}_C = 2{}^p\ell_C(\widehat{{}^p\boldsymbol{\theta}}) - 2\dim(\boldsymbol{\theta}), \tag{2}$$

where $\boldsymbol{\theta}$ is the parameters of the copula C, $\widehat{{}^p\boldsymbol{\theta}}$ is the maximum pseudo-likelihood estimate based on the pseudo-observations ${}^p\mathcal{X}_n$ and ${}^p\ell_C$ is the pseudo-log-likelihood function.

* Corresponding author.
E-mail addresses: lars.jordanger@math.uib.no, lars.jordanger@gmail.com (L.A. Jordanger).

[1] The original formula introduced by Akaike is (−1) times the one given here. Both versions are used in the literature, and the main references for this article use the form given in Eq. (1).

[2] Most statisticians will usually also call this for AIC.

http://dx.doi.org/10.1016/j.spl.2014.06.006
0167-7152/

The use of Eq. (2) has been justified by the belief that in the limit there is a continuous connection between AIC on $\mathfrak{X}_n$ and ${}^p\text{AIC}$ on ${}^p\mathfrak{X}_n$—but it turns out that this is not the case, see Grønneberg and Hjort (2014), and as such one might wonder to what degree results based upon ${}^p\text{AIC}$ should be trusted. A comparison of the small-sample results of ${}^p\text{AIC}$, with those obtained from a cross-validation Copula Information Criterion, has thus been performed in order to see how the bias-correcting terms estimated from the data affects the ranking of the proposed models.

2. The copula information criteria

There are altogether three copula information criteria. The two criteria introduced in Grønneberg and Hjort (2008) are of less practical interest since they are not generally applicable.[3] The xv-CIC from Grønneberg and Hjort (2014) does not suffer from this problem, and can be used as a general model selection tool. A sketch of the arguments leading to xv-CIC will now be presented. (Further details can be found in e.g. Claeskens and Hjort (2008, chapter 2).)

If the observations $\mathfrak{X}_n$ are assumed to originate from some unknown data-generating model F° (with pdf f°), and there is a collection of proposed parametric model families $\{F_k\}$ (with pdfs f_k and parameters $\boldsymbol{\alpha}_k$), then an approach by a Kullback–Leibler information criterion can be used in order to find the optimal approximating model. The argument is based on a minimisation of

$$\text{KLIC}(F^\circ, F) \stackrel{\text{def}}{=} \text{E}_{f^\circ}\left[\log\frac{f^\circ}{f}\right] = \text{E}_{f^\circ}[\log f^\circ] - \text{E}_{f^\circ}[\log f]\,, \tag{3}$$

which can be used to find the best approximating model in the case of a *known* data generating function F°. The idea is that within each proposed parametric model-family F_k there exists an optimal model with parameter configuration $\boldsymbol{\alpha}_k^\circ$ that minimises the "distance" to the true model F°, and the "closest" candidate can then be selected from the resulting set of optimal models.

In a practical setting, with an *unknown* F°, the observed data $\mathfrak{X}_n$ will be used to estimate Eq. (3)—in fact only the term $\text{E}_{f^\circ}[\log f]$ needs to be estimated since the estimate of the term $\text{E}_{f^\circ}[\log f^\circ]$ is common for all the models. When the multivariate empirical distribution function is used as an approximation for F°, the best estimate of $\boldsymbol{\alpha}_k^\circ$ is the one given by the maximum likelihood estimate $\widehat{\boldsymbol{\alpha}}_k$, which thus selects one optimal estimated model from each model-family F_k. Since both F° and $\boldsymbol{\alpha}_k^\circ$ are unknown, some care must be taken before these estimated models are compared against each other—i.e. there is a need for a bias-correcting term.

The argument leading to the bias-correcting term for a proposed parametric family F with optimal parameter configuration $\boldsymbol{\alpha}^\circ$, starts by a Taylor expansion of $\log f$ around $\boldsymbol{\alpha}^\circ$. Then an asymptotic argument is used to get a more convenient expression. In the end, the value assigned to a parametric model F based on n observations $\mathfrak{X}_n$ is given by

$$\text{Estimate of}\left\{\text{E}_{f^\circ}[\log f]\right\} \stackrel{\text{def}}{=} n^{-1}\left(\ell_F(\widehat{\boldsymbol{\alpha}}) - \text{Tr}\left(\widehat{\boldsymbol{J}}^{-1}\widehat{\boldsymbol{K}}\right)\right), \tag{4}$$

in which ℓ_F is the log-likelihood-function, and $\widehat{\boldsymbol{J}}$ and $\widehat{\boldsymbol{K}}$ are estimators of $\boldsymbol{J} = \text{E}_{f^\circ}[\boldsymbol{I}(\boldsymbol{\alpha}^\circ)]$ with $\boldsymbol{I}(\boldsymbol{\alpha}^\circ) = \frac{\partial^2}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^t}\log f\Big|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}^\circ}$ and $\boldsymbol{K} = \text{Var}_{f^\circ}(\boldsymbol{u}(\boldsymbol{\alpha}^\circ))$ with $\boldsymbol{u}(\boldsymbol{\alpha}^\circ) = \frac{\partial}{\partial\boldsymbol{\alpha}}\log f\Big|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}^\circ}$.

Estimation of the matrices $\boldsymbol{J}$ and $\boldsymbol{K}$ requires that all the first and second order partial derivatives of $\log f$ are evaluated at all the observations in $\mathfrak{X}_n$. This requires extra computations, and the resulting bias-correction will experience finite sample variability. However, the computation becomes trivial when the assumption of a correctly specified family is used, i.e. that F° belongs to it. In this case both $\boldsymbol{J}$ and $\boldsymbol{K}$ equals the Fisher information matrix of F°, which implies that $\text{Tr}\left(\boldsymbol{J}^{-1}\boldsymbol{K}\right) = \text{Tr}\left(\boldsymbol{I}\right) = \dim(\boldsymbol{\alpha})$, i.e. the exact bias-correcting term is known without any computations—and one is essentially back to the AIC criterion from Eq. (1). Thus, the AIC formula is a valid selection method when used on *independent* observations $\mathfrak{X}_n$ and under the assumption of a correctly specified model family F. If this assumption is dropped, the estimated bias-correcting term must be computed from the data in accordance with Eq. (4), and in this case the Takeuchi Information Criterion (TIC), Takeuchi (1976), given by $\text{TIC}_F = 2\ell_F(\widehat{\boldsymbol{\alpha}}) - 2\,\text{Tr}\left(\widehat{\boldsymbol{J}}^{-1}\widehat{\boldsymbol{K}}\right)$, should be used instead of AIC.

The precision of asymptotic tools like AIC and TIC will depend on the number of observations, and simulation studies must be performed in order to test their small-sample behaviour. The AIC- and TIC-based rankings of proposed models coincide to a high degree, i.e. the extra computations needed for the empirically estimated correction term in the TIC-formula may not pay off. Moreover, it might also be preferable to avoid the added finite sample variation from the TIC-formula, and in practice it is thus customary to settle for AIC as the chosen model selection method.

Cross-validation is frequently used when the aim is the best predictive model, i.e. one part of the observations is used to estimate the parameters of the model and the remaining part is used to test the fitness of the estimated model. The version of particular interest with regard to the construction of xv-CIC is the *leave-one-out cross-validation*, given by

$$\text{xv}_n = n^{-1}\sum_{i=1}^{n}\log f_{\widehat{\boldsymbol{\alpha}}(i)}\left(\boldsymbol{X}_i\right), \tag{5}$$

[3] Grønneberg explains in his contribution to Kurowicka and Joe (2010), that the theoretic bias-correcting term can attain an infinite expectation when these two selection methods are used on copula models with some tail-dependence.

in which $\widehat{\boldsymbol{\alpha}}(i)$ is the maximum likelihood estimate based on the sample without the ith observation. The following connection with TIC is the reason for the interest in xv_n,

$$\text{TIC} = 2n\,\text{xv}_n + o_p(1)\,, \tag{6}$$

and this motivates the construction in Grønneberg and Hjort (2014) leading to the generally applicable cross-validation Copula Information Criterion, xv-CIC.

Sklar's theorem, Sklar (1959), states that any d-variate cumulative distribution function $F(\boldsymbol{x})$ can be expressed as

$$F(\boldsymbol{x}) = C(\boldsymbol{F}_{\perp}(\boldsymbol{x})), \tag{7}$$

where $C(\boldsymbol{u})$ is a copula on $[0,1]^d$, and $\boldsymbol{F}_{\perp}(\boldsymbol{x}) \stackrel{\text{def}}{=} \left(F_1\left(x_1\right), \ldots, F_d\left(x_d\right)\right)$ is the vector of marginal cdfs. The copula C is unique when the marginals are continuous.

When the focal point of interest is to describe the multivariate dependences in $\mathfrak{X}_n$, the object of interest is the corresponding copula C°. Techniques that increase the chance of finding the correct copula-model, and good estimates of its parameters $\boldsymbol{\theta}$, should thus be used. It turns out that it may often be preferable to simply avoid the use of parametric models for the marginal distributions $\boldsymbol{F}_{\perp}(\boldsymbol{x})$, and instead use the empirical marginal distributions $\boldsymbol{F}_{n,\perp}(\boldsymbol{x}) \stackrel{\text{def}}{=} \left(F_{n,1}(x_1), \ldots, F_{n,d}(x_d)\right)$, where $F_{n,j}(x_j)$ stands for the $\frac{n}{n+1}$-rescaled empirical marginal.[4] This strategy avoids the risk of misspecified marginals, that can severely affect the estimated parameters of the copula.

Based on observations $\mathfrak{X}_n$, from some continuous model $F^\circ(\boldsymbol{x}) = C^\circ(\boldsymbol{F}^\circ_{\perp}(\boldsymbol{x}))$, a vector $\boldsymbol{F}_{n,\perp}(\boldsymbol{x})$ of empirical marginal distributions is constructed—and there are now two sets of points that should be considered,

$${}^u\mathfrak{X}_n \stackrel{\text{def}}{=} \left\{\boldsymbol{F}^\circ_{\perp}(\boldsymbol{x}_i)\right\}\big|_{\boldsymbol{x}_i \in \mathfrak{X}_n} \quad \text{and} \quad {}^p\mathfrak{X}_n \stackrel{\text{def}}{=} \left\{\boldsymbol{F}_{n,\perp}(\boldsymbol{x}_i)\right\}\big|_{\boldsymbol{x}_i \in \mathfrak{X}_n}. \tag{8}$$

When the goal is to find a good approximation to C°, access to the points in ${}^u\mathfrak{X}_n$ would be preferable—since these correspond to a situation with exact knowledge of the models for all the marginal distributions. But the independent set ${}^u\mathfrak{X}_n$ is unattainable when the marginals of F° is unknown, and the set of dependent pseudo-observations in ${}^p\mathfrak{X}_n$ must be used instead. As n increases ${}^p\mathfrak{X}_n$ approaches ${}^u\mathfrak{X}_n$, and it is plausible to expect that minor modifications of the copula model selection techniques from the parametric realm, like Eq. (2), still should provide decent rankings of the copula models considered.

An analysis akin to the one stated after Eq. (3) was first conducted in Grønneberg and Hjort (2008) for the case of a true unknown copula model C° and a collection of proposed copula-models $\{C_i\}$. The replacement of ${}^u\mathfrak{X}_n$ with ${}^p\mathfrak{X}_n$ implies that a Taylor-expansion with regard to the variates $\boldsymbol{u}$ of the copula must also be included in the construction of the bias-correcting term, which results in a more complicated formula than the one given in Eq. (4)—and which perhaps more importantly shows that the ^{p}AIC from Eq. (2) is not a formally valid model selection method in this semiparametric case.

The above approach did not result in a generally applicable model selection method for the semiparametric case, but the approach motivated by the relationship in Eq. (6) between TIC and xv_n did however succeed. Based on this the cross-validation Copula Information Criterion, xv-CIC, for a copula model C is in Grønneberg and Hjort (2014) defined by

$$\text{xv-CIC}_C \stackrel{\text{def}}{=} 2 \cdot {}^p\ell_C\left({}^p\widehat{\boldsymbol{\theta}}\right) - 2 \cdot \left(\widehat{\delta}_c + \widehat{\delta}_m\right), \tag{9}$$

where ${}^p\widehat{\boldsymbol{\theta}}$ and ${}^p\ell_C$ are as described after Eq. (2), and where the two bias-correcting terms $\widehat{\delta}_c$ and $\widehat{\delta}_m$ both depend on ${}^p\widehat{\boldsymbol{\theta}}$ and account for different effects due to Taylor expansions. A more detailed presentation of these two bias-correcting components is beyond the scope of the present article, whose aim instead is to present some results based upon the first author's master's thesis,[5] where the small-sample adequacy of xv-CIC, which is more asymptotically correct but harder to compute, were investigated and compared to ^{p}AIC, which may not be quite correct but much easier to compute. To our knowledge such a simulation study has not been undertaken before.

3. Setup for simulations

In order to examine how xv-CIC fares as a selection method in the finite sample realm, it must be tested on data with known origin. The main part of this presentation includes data originating from the following five bi-variate one-parameter copula families: `clayton`, `frank`, `gumbel`, `normal` and `t (df=4)`. Of these, the first three are Archimedean copulas, with `gumbel` being an extreme-value copula. The two copula models `normal` and `t` refer respectively to the copulas *corresponding* to the bivariate normal distribution and the bivariate student's t-distribution.

Using the R-package copula, Ivan Kojadinovic and Jun Yan (2010), samples ${}^u\mathfrak{X}_n$ of size $N \in \{100, 250, 500, 1000\}$ were drawn from all of the five copula models, with parameter values chosen to give values of the rank correlation coefficient

[4] The rescaling ensures that points on the edge of the unit (hyper)cube in $\mathbb{R}^d$ is avoided, which is important since many copula models of interest have heavy tail-dependence, and points on the boundary could then introduce infinities into the calculations.

[5] Semiparametric Model Selection for Copulas, http://hdl.handle.net/1956/6778.

Table 1
xv-CIC versus ${}^{p}AIC$, $N = 250$ and $\tau = 0.5$—counting.

d.cop	IC	clayton	frank	gumbel	normal	t (df=4)
clayton	${}^{p}AIC$	4992	2	0	5	1
clayton	xv-CIC	4974	9	0	11	6
frank	${}^{p}AIC$	3	4663	28	270	36
frank	xv-CIC	0	4741	34	201	24
gumbel	${}^{p}AIC$	0	28	4595	170	207
gumbel	xv-CIC	0	36	4754	98	112
normal	${}^{p}AIC$	8	133	229	4332	298
normal	xv-CIC	1	196	362	4208	233
t (df=4)	${}^{p}AIC$	8	27	275	186	4504
t (df=4)	xv-CIC	4	42	431	211	4312

Table 2
Coincidence of ${}^{p}AIC$ and xv-CIC, with 95% confidence interval.

N	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	All
100	90.260 ± 0.1172	91.264 ± 0.1178	92.540 ± 0.1186	91.354 ± 0.0681
250	95.016 ± 0.1202	96.912 ± 0.1214	97.812 ± 0.1219	96.580 ± 0.0700
500	98.604 ± 0.1224	99.380 ± 0.1229	99.732 ± 0.1231	99.238 ± 0.0709
1000	99.608 ± 0.1230	99.964 ± 0.1233	99.976 ± 0.1233	99.849 ± 0.0711

Table 3
Hit-rates for ${}^{p}AIC$ and xv-CIC, the case $N = 100$ and $\tau = 0.25$.

IC	clayton	frank	gumbel	normal	t (df=4)
${}^{p}AIC$	81.28 ± 1.081	47.72 ± 1.384	64.22 ± 1.329	37.84 ± 1.344	63.80 ± 1.332
xv-CIC	73.34 ± 1.225	51.32 ± 1.385	68.96 ± 1.282	37.82 ± 1.344	61.10 ± 1.351

Kendall's τ in the set $\{0.25, 0.5, 0.75\}$. Since the pseudo-observations ${}^{p}\mathfrak{X}_n$ are created by the help of the empirical marginals, it is sufficient to work directly upon the samples ${}^{u}\mathfrak{X}_n$ generated from the copula models, cf. Fermanian et al. (2004, Lemma 1). A total of 5000 samples were created for each combination of copula, N and τ—and then all of the five copula models were fitted to these samples and ranked according to values of ^{p}AIC and xv-CIC.

4. Results

Table 1 is representative for the behaviour of ^{p}AIC and xv-CIC for all the twelve combinations of N and τ under consideration. The first column "d.cop" specifies the *data-generating copula* that the sample originated from, while the second column "IC" contains the *information criterion* that filled in the remaining five columns according to how often the selection criteria picked the proposed copula as the model for the sample at hand. It seems like the two selection methods follow each other rather closely, both with regard to correct and erroneous proposals, without one being superior.

The coincidence-percentages, i.e. the fraction of times that ^{p}AIC and xv-CIC proposed the same model regardless of whether or not it was the true model, is given in Table 2. The estimated 95% confidence intervals are based upon the asymptotic approximation to the standard normal distribution, which can be used due to the size of the data-sets (25 000 for each cell in the τ-columns, 75 000 for the all-column).

These *coincidence*-percentages reveal that ^{p}AIC and xv-CIC necessarily also must have close *returned-the-true-model*-percentages (hit-rates)—and that the sample size must be small in order for one of them to clearly outperform the other. Since low values of τ imply that the data-generating copula-models all are closer to the independence copula, and since the transformation ${}^{u}\mathfrak{X}_n \to {}^{p}\mathfrak{X}_n$ induces more randomness for low values of N, the most distinct differences between ^{p}AIC and xv-CIC are found for the samples with $\tau = 0.25$ and $N = 100$. Table 3 gives the hit-rates for this case, with 95% confidence intervals. There is no indication that one of them is superior—and similar tables for the other combinations of τ and N support that impression.

Some of the hit-rates in Table 3 are rather low, but the improvement when the sample-size increases is reassuring. For $N = 500$ and $N = 1000$ the lowest observed hit-rates were respectively 83.76% and 95.02%, whereas the highest in both cases turned out to be 100%.

The true data-generating copula-models, and the samples ${}^{u}\mathfrak{X}_n$, will in practice be unknown. If a model selection strategy based directly on ${}^{u}\mathfrak{X}_n$ is compared against ^{p}AIC and xv-CIC used on ${}^{p}\mathfrak{X}_n$, that should give an idea to what extent the extra randomness from the transformation ${}^{u}\mathfrak{X}_n \to {}^{p}\mathfrak{X}_n$ affected the conclusion, or to what extent the true model differs from the finite sample Kullback–Leibler best choice. In order to investigate this, a bootstrap estimate for Kullback–Leibler information for model selection was used, i.e. the WIC-bootstrap, based on 1000 bootstrap-replicates, on the case $N = 100$, $\tau = 0.25$. See Shibata (1997) for a description of the WIC-method and further references.

Table 4
Coincidence of $\mathring{A}IC$ and xv-CIC with WIC-bootstrap, $N = 100$ and $\tau = 0.25$.

IC	clayton	frank	gumbel	normal	t (df=4)
$\mathring{A}IC$	86.86 ± 0.936	73.58 ± 1.222	79.50 ± 1.119	71.38 ± 1.253	81.22 ± 1.082
xv-CIC	83.60 ± 1.026	74.88 ± 1.202	80.34 ± 1.101	70.98 ± 1.258	79.74 ± 1.114

The coincidence percentages for ^{p}AIC and xv-CIC against the WIC-selection turned out to be respectively 78.51 ± 0.509 and 77.91 ± 0.514, i.e. even for the small sample size of $N = 100$ they are almost close enough to reject a hypothesis that they should be considered as truly different selection techniques. Moreover, when the knowledge of the data-generating models is used to stratify the results, see Table 4,[6] the same affinity/aversion for the different models shows up as those present in Table 3.

From Tables 3 and 4 it seems like the two selection criteria have some affinity for different models, which could be related to the behaviour of the bias-correcting term in the xv-CIC-formula. The plots in Fig. 1 show how the distribution of this term varies for two N-τ-combinations when fitting the copula `t (df=4)`, where the knowledge of the true data-generating copula-models has been used to stratify the observed bias-corrections into groups of 5000 observations. The dotted line at $x = 1$ shows the simple bias-correction used by ^{p}AIC. These plots indicate that the bias-correction in addition to its inherent variation with the model fitted also varies quite a bit with different values of τ and N and different data-generating models.

The increased span of bias-corrections for the $N = 1000$-case originates from the fact that the xv-CIC-formula does not rescale the bias-correction with the size of the sample, cf. Eq. (4). The tenfold increase in sample-size thus represents a significantly improved precision, which might be a reason for the improved coincidence-percentages observed in Table 2. That table says that the bias-correction in most cases did not alter the ranking based on the values of the pseudo-log-likelihood ${}^{p}\ell$,[7] and the hit-rates from Table 3 indicate that it hardly was any improvement in the cases where the rankings were affected.

Conclusion

For the five bivariate one-parameter copula models considered in this section, the difference in performance between xv-CIC and ^{p}AIC is altogether rather small. When the sample-size is large enough for them to perform good, their difference is small enough to make it tempting to consider them to be interchangeable as selection techniques. This implies that there has been no payoff for the extra computational investment that was made in the production of the xv-CIC-values, and it thus seems reasonable to propose the same practice here as the one used in the fully parametric setting—where the "costly" computation of TIC without further ado is replaced with the "inexpensive" AIC.

5. A C-vine example in three dimensions

The previous section only considered bi-variate one parameter copula-models, and the question remains if the conclusion would differ if xv-CIC were tested against ^{p}AIC for more complicated cases. However, the simultaneous estimation procedure needed for an investigation of how xv-CIC fares on such models quickly becomes computationally intractable. A more modest task has thus been performed, investigating the effect of using xv-CIC instead of AIC as the selection tool for the (conditional) bivariate copula models of a C-vine. The effect on the AIC-rankings when observations ${}^{u}\mathfrak{X}_n$ are replaced with pseudo-observations ${}^{p}\mathfrak{X}_n$ has also been included in Fig. 2 (explained below), where data from 729 different three dimensional C-vines are presented.

Observations ${}^{u}\mathfrak{X}_n$ with more than two variates can be modelled by vine copulas and the pairwise copula construction—i.e. a d-variate copula model can be approximated by first selecting a vine-copula structure based on the data, before bivariate copula models are selected for the different levels of this vine, see e.g. Aas et al. (2009), Brechmann (2010), Haff (2012) and references therein. For the case of a three dimensional copula model, this approximation requires a simple C-vine and the specification of three bivariate copula models (of which one is a conditional copula).

To keep things computationally tractable, the bi-variate copula models were restricted to the three one-parameter copula-models `clayton`, `gumbel` and `normal`, with values for Kendall's τ as before—which for the case of a three dimensional C-vine gave 729 different data-generating models. A total of 200 samples ${}^{u}\mathfrak{X}_n$ of size $N = 500$ was drawn from each model using functions from the R-package VineCopula, Schepsmeier et al. (2013), and the corresponding pseudo-observations ${}^{p}\mathfrak{X}_n$ were then created. The model-estimating function in the VineCopula-package was used on both ${}^{u}\mathfrak{X}_n$ and ${}^{p}\mathfrak{X}_n$, before a modified xv-CIC-version was used on ${}^{p}\mathfrak{X}_n$.

[6] The coincidence-percentage *between* the bootstrap-based selection technique and ordinary AIC, both used on samples ${}^{u}\mathfrak{X}_n$, turned out to be 95.02 ± 0.269, and their hit-rates (against the true models) were almost identical, i.e. 62.98 ± 0.598 and 62.99 ± 0.598, so Table 4 would have looked almost the same if the comparison had been against the models proposed by ordinary AIC instead. When the AIC-comparison is made for the higher values of N, the models obtained from an analysis based on ${}^{u}\mathfrak{X}_n$ and those based on ${}^{p}\mathfrak{X}_n$ coincide almost completely.

[7] Since all the models have one parameter, the ^{p}AIC-ranking coincides with the ${}^{p}\ell$-ranking.

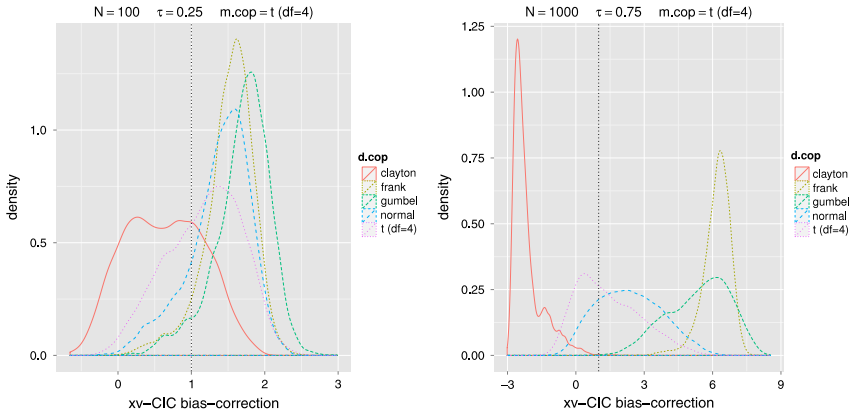

Fig. 1. Stratified behaviour of bias-correction for xv-CIC, when fitting `t (df=4)`.

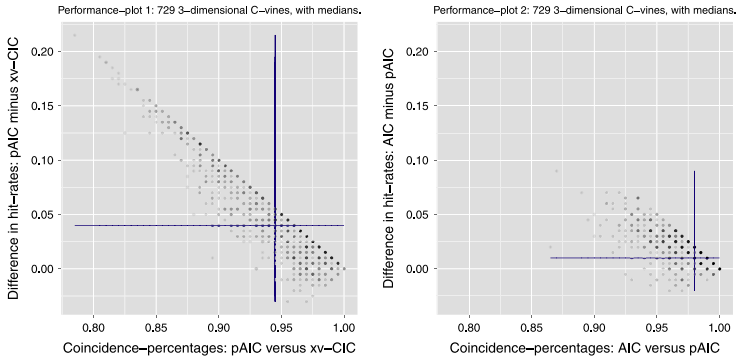

Fig. 2. Coincidence-percentages and hit-rates for the simple three dimensional case.

The left plot in Fig. 2 compares the performance of the xv-CIC-adjusted selection method with the ordinary AIC-selection method used on ${}^{p}\mathcal{X}_n$. The coincidence-percentages are along the horizontal axis whereas the vertical axis presents the difference in hit-rates. There is some overlap among the points, and lines representing the two medians have been included. The impression is that there is no gain in hit-rate from an approach using the xv-CIC-modified selection, in particular since the low coincidence-percentage cases had better hit-rates for the ordinary ^{p}AIC selection strategy.

The right plot does not consider the xv-CIC at all, but instead gives an impression of how the transformation ${}^{u}\mathcal{X}_n \to {}^{p}\mathcal{X}_n$ affected the selection process in this case with samples of size $N = 500$. For decent-sized data-sets, it seems safe to assume that the result from a ^{p}AIC-selection based on the pseudo-observations ${}^{p}\mathcal{X}_n$ should return more or less the same result as an AIC-selection based upon ${}^{u}\mathcal{X}_n$.

This simple example indicates that AIC used on pseudo-observations ${}^{p}\mathcal{X}_n$ does a good job, and that there is little/no gain from the use of xv-CIC. Moreover, the xv-CIC-modified selection method for the bivariate copulas in the C-vine could actually be less straightforward to use, since the asymptotic validity of xv-CIC has not been established for the *conditional* bivariate copulas needed for the higher levels of the vine.

6. Some further comments

The present analysis does not cover the case of copula models with more than one parameter. The increased computational cost incurred in the computation of xv-CIC for copula models with several parameters implies that ^{p}AIC probably still would be the standard alternative even if xv-CIC should happen to be a slightly better selection method in some cases.

Regardless of whether xv-CIC or ^{p}AIC is used, one needs to keep in mind that these selection methods do not give any information about the quality of the chosen model. In particular, if there is no adequate models among those copula attempted fit to the data, then the same will be true for the selected approximation to the least false of them. It is thus important to apply goodness-of-fit tests to the chosen models, in order to see if they may be trusted or not, see e.g. Berentsen et al. (2014).

The purpose of this work was to do a comparison of xv-CIC and ^{p}AIC, and readers interested in model selection for copula models in general can consult e.g. Aas et al. (2009), Dißmann et al. (2013).

Acknowledgements

The authors are most grateful to the anonymous referees who provided several valuable comments and suggestions.

References

Aas, K., Czado, C., Frigessi, A., Bakken, H., 2009. Pair-copula constructions of multiple dependence. Insurance Math. Econom. 44, 182–198.
Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automat. Control 19, 716–723.
Berentsen, G.D., Støve, B., Tjøstheim, D., Nordbø, T., 2014. Recognizing and visualizing copulas: an approach using local Gaussian approximation. Insurance Math. Econom. 57, 90–103.
Brechmann, E.C., 2010. Truncated and simplified regular vines and their applications. Master's thesis, Technische Universität München. URL http://mediatum.ub.tum.de/doc/1079285/1079285.pdf.
Claeskens, G., Hjort, N.L., 2008. Model Selection And Model Averaging. Cambridge University Press, Cambridge.
Dißmann, J., Brechmann, E.C., Czado, C., Kurowicka, D., 2013. Selecting and estimating regular vine copulae and application to financial returns. Comput. Statist. Data Anal. 59, 52–69. URL http://www.sciencedirect.com/science/article/pii/S0167947312003131.
Fermanian, J.-D., Radulovic, D., Wegkamp, M., 2004. Weak convergence of empirical copula processes. Bernoulli 10 (5), 847–860. URL http://dx.doi.org/10.3150/bj/1099579158.
Grønneberg, S., Hjort, N.L., 2008. The copula information criterion. Technical Report 7, Department of Matehmatics, University of Oslo.
Grønneberg, S., Hjort, N. L., 2014. The copula information criteria. Scand. J. Stat. URL http://dx.doi.org/10.1111/sjos.12042.
Haff, I.H., 2012. Pair-copula constructions—an inferential perspective (Ph.D. thesis). University of Oslo, Faculty of Mathematics and Natural Sciences.
Ivan, Kojadinovic, Jun, Yan, 2010. Modeling multivariate distributions with continuous margins using the copula R package. J. Stat. Softw. 34 (9), 1–20. URL http://www.jstatsoft.org/v34/i09/.
Kurowicka, D., Joe, H., 2010. Dependence Modeling—Vine Copula Handbook. World Scientific, Singapore.
Schepsmeier, U., Stoeber, J., Brechmann, E.C., Gräler, B., 2013. VineCopula: statistical inference of vine copulas. R package version 1.2. URL http://CRAN.R-project.org/package=VineCopula.
Shibata, R., 1997. Bootstrap estimate of Kullback–Leibler information for model selection. Statist. Sinica 7, 375–394.
Sklar, A., 1959. Fonctions de répartition à n dimensions et leurs marges. Publ. Inst. Stat. Univ. Paris 8, 229–231.
Takeuchi, K., 1976. Distribution of informational statistics and a criterion of model fitting. Suri-Kagaku (Math. Sci.) 153, 12–18 (in Japanese).