JOURNAL OF TIME SERIES ANALYSIS
J. Time Ser. Anal. **41**: 210–228 (2020)
Published online 08 August 2019 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12500

ORIGINAL ARTICLE

ESTIMATING THE MEAN DIRECTION OF STRONGLY DEPENDENT CIRCULAR TIME SERIES

JAN BERAN[a] AND SUCHARITA GHOSH[b]

[a] *Department of Mathematics and Statistics, University of Konstanz, Konstanz, Germany*
[b] *Statistics Lab, Swiss Federal Research Institute WSL, Birmensdorf, Switzerland*

A class of circular processes based on Gaussian subordination is introduced. This allows for flexible modelling of directional time series with long-range dependence. Based on limit theorems for subordinated processes and consistent estimation of nuisance parameters, asymptotic confidence intervals for the mean direction are derived. Extensions to cases where the direction depends on explanatory variables are also considered. Simulations and a data example illustrate the proposed method.

Received 02 November 2018; Accepted 06 July 2019

Keywords: Circular time series; mean direction; long-range dependence; Gaussian subordination; confidence interval

JEL Codes C22; C49

MOS subject classifications: 62M10; 62M09; 62G20.

1. INTRODUCTION

Directional or circular data arise in many scientific fields such as meteorology, oceanography, biology, neuroscience, bioinformatics, geoscience and cosmology. Often observations occur in a temporal sequence. Given a circular time series $\vartheta_1, \ldots, \vartheta_n \in [0, 2\pi)$ one of the first questions is inference about the mean direction μ defined by

$$E\left[\exp(i\vartheta)\right] = R\exp(i\mu), \tag{1}$$

where $R > 0$. Here, it is assumed that ϑ and $\vartheta_1, \ldots, \vartheta_n$ all have the same (marginal) distribution on $[0, 2\pi)$. In this article, we consider inference about μ for circular time series that exhibit long-range dependence. For i.i.d. observations Fisher and Lewis (1983) introduced simple asymptotic confidence intervals for μ, based on the central limit theorem for

$$\left(\bar{C}, \bar{S}\right) = \left(n^{-1}\sum_{j=1}^{n}\cos\vartheta_j, n^{-1}\sum_{j=1}^{n}\sin\vartheta_j\right).$$

For weakly dependent observations a similar central limit theorem holds. However, in some applications temporal dependence is much stronger such that $\bar{S}$, and possibly also $\bar{C}$, has a slower rate of convergence than under independence or short-range dependence. An example is shown in Figures 1 and 2. Figure 2(d) displays the periodogram (in log–log coordinates) of $\sin(\vartheta_j - \hat{\mu})$ where ϑ_j are daily wind directions observed in Milwaukee between 30 January and 31 December 2017, and $\hat{\mu}$ is the estimated mean direction. The negative slope in the log–log plot indicates a hyperbolic pole of the spectral density at the origin, and thus long-range dependence.

* Correspondence to: Jan Beran, Department of Mathematics and Statistics, University of Konstanz, Universitaetsstrasse 10, 78457 Konstanz, Germany.
E-mail: jan.beran@uni-konstanz.de

© 2019 John Wiley & Sons Ltd

Konstanzer Online-Publikations-System (KOPS)
URL: http://nbn-resolving.de/urn:nbn:de:bsz:352-2-qn2sncvrlmh14

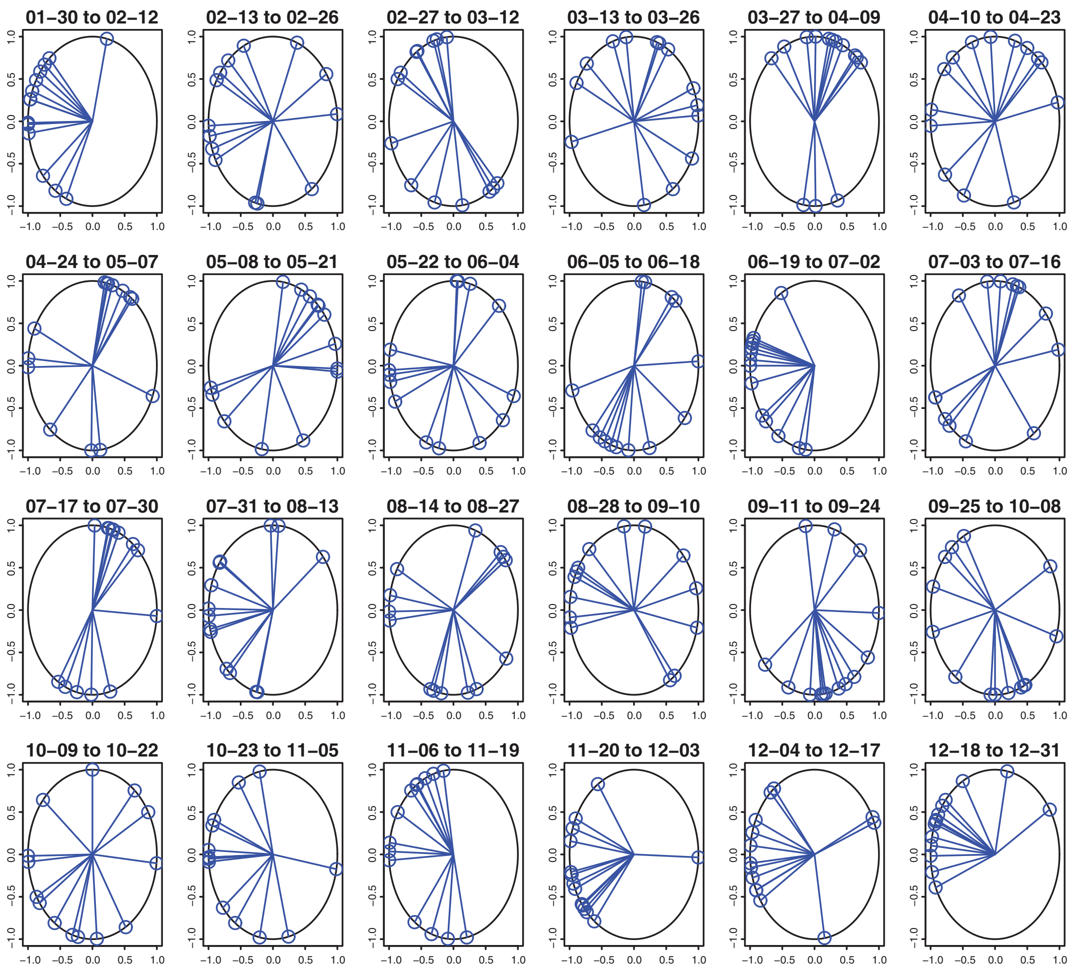

Figure 1. Milwaukee daily average wind directions between 30 January and 31 December, 2017. Here ϑ_j ($j = 1, \ldots, n$) are visualized as points on the unit circle, connected to the origin. Each plot corresponds to a 2-week periods, that is 14 consecutive observations [Color figure can be viewed at wileyonlinelibrary.com]

In this article, a class of circular processes based on Gaussian subordination is introduced. The approach allows for flexible modelling of circular time series with long-range dependence and arbitrary marginal circular distributions. Limit theorems for $\bar{C}$ and $\bar{S}$ and consistent estimation of nuisance parameters are then combined to obtain asymptotic confidence intervals for μ. The results are generalized to non-stationary circular processes where the mean direction depends on explanatory variables. Simulations and a data example illustrate the proposed methods.

2. STRONGLY DEPENDENT CIRCULAR TIME SERIES

There is an extended literature on circular data. Excellent books on the topic are for instance Mardia (1972), Fisher (1993), Mardia and Jupp (1999), Jammalamadaka and SenGupta (2001), Pewsey and Neuhauser (2013) and Ley and Verdebout (2017). Most methods have been developed for i.i.d. observations. Circular processes and circular time series analysis are considered for example in Wehrly and Johnson (1980), Breckling (1989), Fisher and Lee (1994), Kato (2010), Modlin *et al.* (2012) and Wang and Gelfand (2014). For stationary circular time

 © 2019 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12500

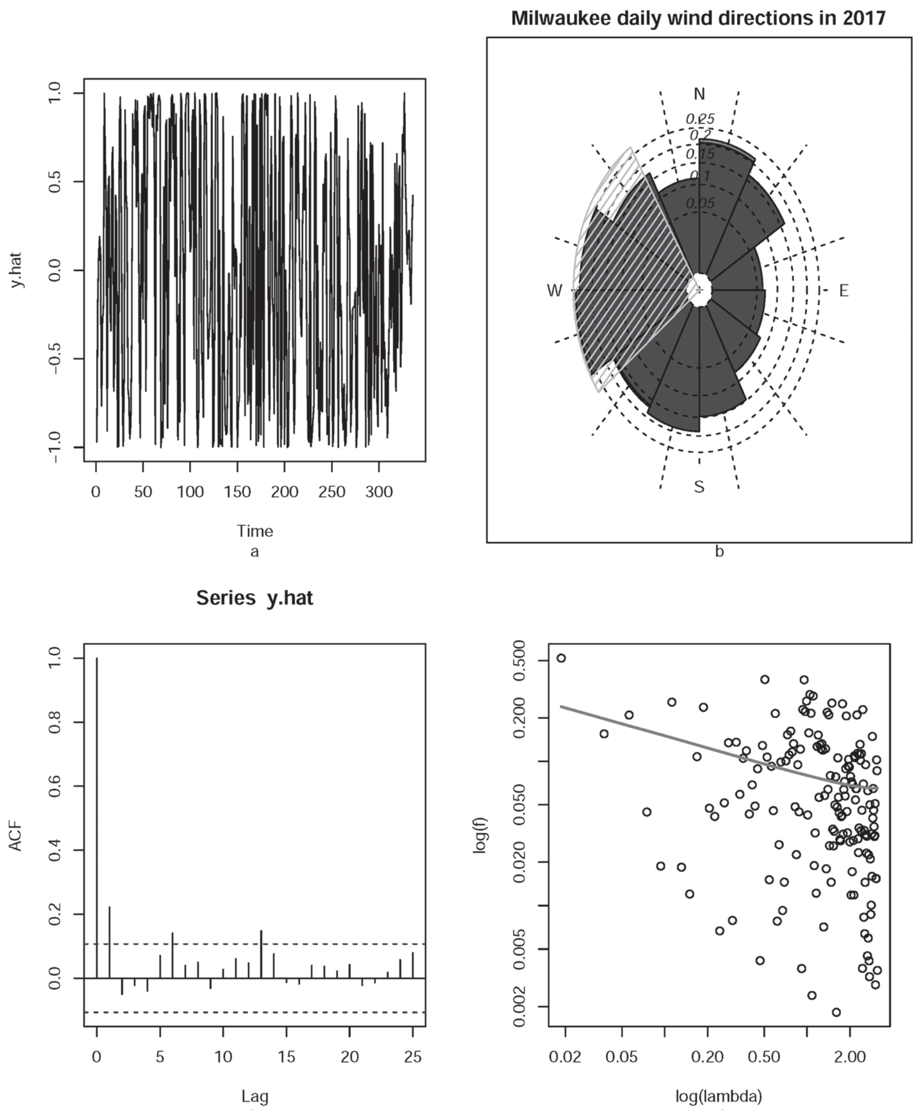

Figure 2. Milwaukee daily wind directions: time series plot of $\hat{Y}_j = \sin\left(\vartheta_j - \hat{\mu}\right)$ (a); windrose plot and an approximate 95%-confidence interval for μ based on (13) (b); empirical autocorrelation function of $\hat{Y}_j$ (c); and periodogram of $\hat{Y}_j$ (in log–log-coordinates) together with the fitted spectral density (d)

series, autocorrelation can be defined for instance by

$$\rho_{\text{circular}}(k) = \frac{E\left[\sin\left(\vartheta_j - \mu\right)\sin\left(\vartheta_{j+k} - \mu\right)\right]}{E\left[\sin^2\left(\vartheta_1 - \mu\right)\right]} \tag{2}$$

(Jammalamadaka and Sarma, 1988). With the exception of Di Marzio *et al.* (2012), circular time series models discussed in the literature are weakly dependent. In contrast, here we will define circular time series with long-range dependence (or strong dependence). In the context of real valued second-order stationary time series $Z_j \in \mathbb{R}$ ($j \in \mathbb{Z}$) with autocovariance function $\gamma_Z(k) = cov(Z_j, Z_{j+k})$, Z_j is said to exhibit long-range (or strong) dependence,

 © 2019 John Wiley & Sons Ltd *J. Time Ser. Anal.* **41**: 210–228 (2020)
DOI: 10.1111/jtsa.12500

if $\sum \gamma_Z(k) = \infty$. More specifically, it is often assumed that $\gamma_Z(k) \sim c_{\gamma,Z} k^{2d-1}$ $(k \to \infty)$, or that the spectral density has a pole at the origin characterized by

$$f_Z(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_Z(k) e^{-ik\lambda} \underset{\lambda \to 0}{\sim} c_{f,Z} |\lambda|^{-2d} \tag{3}$$

with $d \in (0, \frac{1}{2})$ and $0 < c_{\gamma,Z}, c_{f,Z} < \infty$. Here, '$\sim$' means that the ratio of the left- and right-hand side converges to 1. For references to the extended literature on long-memory processes see for instance Beran (1994), Giraitis *et al.* (2012) and Beran *et al.* (2013).

To obtain a circular time series with long-range dependence, we apply the method of Gaussian subordination. Thus, let $(Z_j, j \in \mathbb{Z})$ be a stationary Gaussian process with $E(Z_j) = 0$, $var(Z_j) = 1$, autocovariance function γ_Z and spectral density f_Z such that (3) holds. Moreover, let G be an absolutely continuous circular distribution function with density $g = G'$, and denote by Φ and $\varphi = \Phi'$ the standard normal distribution and density function respectively. Recall that a circular distribution is a probability distribution of a random variable whose values are angles, usually taken to be in the range $[0, 2\pi)$. We assume that the observed circular process ϑ_j is defined by

$$\vartheta_j = G^{-1}\left(\Phi\left(Z_j\right)\right) \; (j \in \mathbb{Z}). \tag{4}$$

More generally, we may define $\vartheta_j = \Psi\left(Z_j\right)$ where $\Psi : \mathbb{R} \to [0, 2\pi)$ is a Lebesgue measurable function. A time series of this type is said to be subordinated to the Gaussian process Z_j (see e.g. Rosenblatt, 1961, 1979; Taqqu 1975, 1979; Dobrushin and Major, 1979; Dobrushin, 1980). Note that the more specific definition (4) is useful for establishing an explicit link to prespecified circular distribution functions.

For Gaussian subordination models, Hermite polynomials, defined by

$$H_q(z) = (-1)^q \exp\left(\frac{1}{2}z^2\right) \frac{d^q}{dz^q} \exp\left(-\frac{1}{2}z^2\right) \; (q \in \mathbb{N})$$

play an essential role. In the L^2-space of real valued functions H with $\|H\|^2 = \int |H(z)|^2 \varphi(z) \mathrm{d}z < \infty$, H_q $(q = 0, 1, \ldots)$ build an orthogonal basis. In particular, we have an L^2-representation

$$H(z) = E[H(Z)] + \sum_{q=1}^{\infty} \frac{a_q}{q!} H_q(z),$$

where Z is a standard normal random variable, and

$$a_q = \left\langle H, H_q \right\rangle = \int_{-\infty}^{\infty} H(z) H_q(z) \varphi(z) \mathrm{d}z$$

is the qth Hermite coefficient. A function H is called of Hermite rank m, if $a_m \neq 0$ and $a_j = 0$ $(j < m)$ (Taqqu, 1975).

The long-memory properties of Z_j are inherited by ϑ_j in the following sense.

Lemma 1. Let ϑ_j $(j \in \mathbb{Z})$ be defined by (4). Suppose furthermore that the Hermite rank of $H(z) = \sin\{G^{-1}(\Phi(z)) - \mu\}$ is $m \geq 1$ and $d > \frac{1}{2}(1 - m^{-1})$. Then

$$\rho_{\text{circular}}(k) \underset{k \to \infty}{\sim} c_m k^{2d_m - 1} \tag{5}$$

with

$$d_m = \frac{1}{2}(1 + (2d - 1)m) \in \left(0, \frac{1}{2}\right),$$
$$c_m = \frac{a_m^2 c_{\gamma,Z}^m}{m! E\left[\sin^2\left(\vartheta_1 - \mu\right)\right]},$$

and

$$a_m = \int H_m(z) \sin\{G^{-1}(\Phi(z)) - \mu\} \varphi(z)\, \mathrm{d}z.$$

3. ASYMPTOTIC CONFIDENCE INTERVALS FOR THE MEAN DIRECTION

3.1. Confidence Intervals With Known Nuisance Parameters

The mean direction μ is defined by (1). A standard estimator of μ is defined by

$$\exp(i\hat{\mu}) = \bar{R}^{-1}\left(\bar{C} + i\bar{S}\right), \tag{6}$$

where $\bar{R}^2 = \bar{C}^2 + \bar{S}^2$. Simple asymptotic confidence intervals for μ are derived by Fisher and Lewis (1983) as follows. Let $\vartheta^* = \vartheta - \mu$, $E[\exp(2i\vartheta^*)] = \alpha_2 + i\beta_2$, $\bar{C}^* = n^{-1}\sum \cos \vartheta_j^*$ and $\bar{S}^* = n^{-1}\sum \sin \vartheta_j^*$. Then $E[\exp(i\vartheta^*)] = \exp(-i\mu)E[\exp(i\vartheta)] = R$, $\bar{C}^* = \bar{R}\cos(\hat{\mu} - \mu)$, $\bar{S}^* = \bar{R}\sin(\hat{\mu} - \mu)$, $E(\cos \vartheta^*) = R$, $E(\sin \vartheta^*) = 0$, $E(\cos 2\vartheta^*) = \alpha_2$, $E(\sin 2\vartheta^*) = \beta_2$, $var(\cos \vartheta^*) = \frac{1}{2}(1 + \alpha_2 - 2R^2)$, $var(\sin \vartheta^*) = \frac{1}{2}(1 - \alpha_2)$, and $\bar{C}^{*2} + \bar{S}^{*2} = \bar{C}^2 + \bar{S}^2 = \bar{R}^2$. Suppose now that ϑ_j are i.i.d. Then

$$\sqrt{n}\frac{\bar{R}\sin(\hat{\mu} - \mu)}{\sqrt{\frac{1}{2}\left(1 - \alpha_2\right)}} \underset{d}{\rightarrow} Z, \tag{7}$$

where Z is a standard normal variable. Thus, noting that

$$\lim_{n\to\infty} P\left(\left|\bar{R}\sin(\hat{\mu} - \mu)\right| \leq n^{-\frac{1}{2}}\sqrt{\frac{1}{2}\left(1 - \alpha_2\right)} z_{1-\alpha/2}\right) = 1 - \alpha. \tag{8}$$

an approximate confidence interval for μ, at confidence level $(1 - \alpha)$, is given by

$$\hat{\mu} \pm \arcsin\left(\sqrt{\frac{\frac{1}{2}\left(1 - \alpha_2\right)}{n\bar{R}^2}} z_{1-\alpha/2}\right). \tag{9}$$

Here, $z_{1-\alpha/2}$ denotes a $(1 - \alpha/2)$-quantile of the standard normal distribution. To estimate the unknown nuisance parameter α_2, Fisher and Lewis suggest $\hat{\alpha}_2 = n^{-1}\sum \cos(2(\vartheta_j - \hat{\mu}))$. Note that the length of confidence intervals defined by (9) is at most π. For reasonable sample sizes, and asymptotically, this restriction does not matter. Note also that in principle, in particular for very small sample sizes, the argument of the arcsine function could be outside the range $[-1, 1]$. In such cases, the asymptotic formula given in (9) is not applicable, and a larger sample size would be needed.

 © 2019 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12500

Under the assumptions of Lemma 1, (7) no longer holds. Instead we have:

Theorem 1. Let ϑ_j ($j \in \mathbb{Z}$) be defined by (4), the process Z_j has spectral density f_Z such that (3) holds, and denote by

$$a_1 = \int_{-\infty}^{\infty} zH(z)\,\varphi(z)\,\mathrm{d}z$$

the first Hermite coefficient of

$$H(z) = \sin\left\{G^{-1}(\Phi(z)) - \mu\right\}.$$

Suppose that $a_1 \neq 0$. Then

$$n^{\frac{1}{2}-d}\bar{S}^* \underset{d}{\to} a_1\sqrt{\nu(d)\,c_{f,Z}}Z,$$

where $Z \sim \mathrm{N}(0, 1)$ and

$$\nu(d) = \frac{2\sin\pi d}{d(2d+1)}\Gamma(1-2d)\left(d \in \left(0, \frac{1}{2}\right)\right).$$

Theorem 1 implies

$$\lim_{n\to\infty} P\left(\left|\bar{R}\sin(\hat{\mu}-\mu)\right| \leq a_1 n^{d-1/2}\sqrt{\nu(d)\,c_{f,Z}}z_{1-\alpha/2}\right) = 1-\alpha.$$

Therefore (9) has to be replaced by

$$\hat{\mu} \pm \arcsin\left(a_1 n^{d-1/2}\sqrt{\nu(d)\,c_{f,Z}}z_{1-\alpha/2}\right). \tag{10}$$

Remark 1. If $a_1 \neq 0$, then the length of confidence intervals given by (10) converges to zero at a slower rate than under independence or short-range dependence. In cases where $a_1 = 0$, the Hermite rank of H is at least two, and a faster rate of convergence is obtained. However, the limiting distribution is no longer Gaussian and difficult to compute in general.

Remark 2. The assumption of Hermite rank one is reasonable in most situations. As an example, consider the von Mises and the wrapped normal distributions (see Section 4 for definitions). Figure 3 shows a_1 for von Mises distributions with scale parameter $0.05 \leq \kappa \leq 6$ (full black line). Note that it is sufficient to vary κ, since a_1 does not depend on μ. These results are compared to the first Hermite coefficient of $\sin\{G^{-1}(\Phi(z)) - \mu\}$ for wrapped normal distributions with $\sigma^2 = 1/\kappa$ (dotted red line). The reason for matching σ^2 this way is that, for $\kappa \to \infty$, the von Mises distribution approximates the wrapped normal distribution with $\sigma^2 = 1/\kappa$ (see e.g. Mardia and Jupp, 1999). Figure 3 shows that, for both distribution families, a_1 is far from zero for all values of κ.

Remark 3. Note that (10) is asymptotically equivalent to

$$\hat{\mu} \pm \arcsin\left(a_1 n^{-1/2}\sqrt{\nu(d) f_Z(n^{-1})}z_{1-\alpha/2}\right). \tag{11}$$

Setting $\nu(0) = \lim_{d\to 0}\nu(d) = 2\pi$, (11) is also applicable to processes defined by (4) with $d = 0$. In particular, for i.i.d observations (11) is asymptotically equivalent to the formula obtained by Fisher and Lewis (1983) (Equation (9)).

 © 2019 John Wiley & Sons Ltd

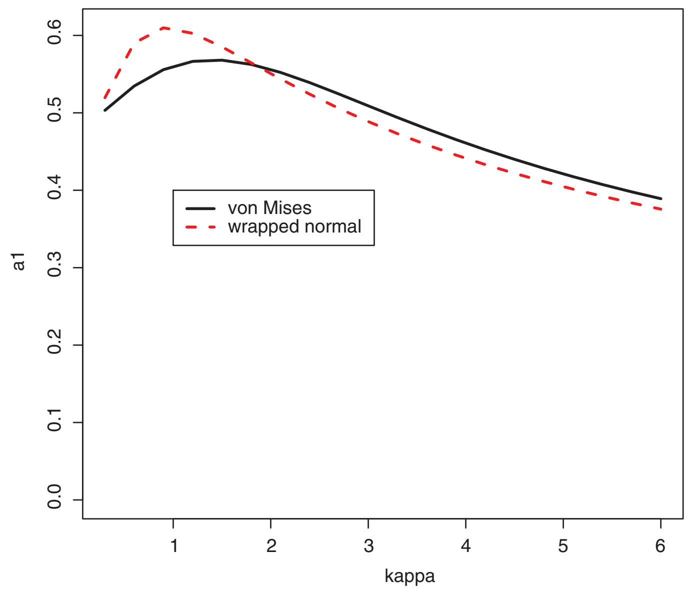

Figure 3. First Hermite coefficient a_1 of $\sin\{G^{-1}(\Phi(Z)) - \mu\}$ for von Mises distributions with $0.05 \le \kappa \le 6$ (full black line) and wrapped normal distributions with $\sigma^2 = 1/\kappa$ (dotted red line) [Color figure can be viewed at wileyonlinelibrary.com]

3.2. Confidence Intervals With Unknown Nuisance Parameters

In practice, formula (10) cannot be used directly, because it includes the unknown nuisance parameters $c_{f,Z}$, d and a_1. Note in particular that Z_j is an unobserved latent process and the transformation in (4) is unknown. To obtain data based confidence intervals, we note first that the autocovariance function and the spectral density of

$$Y_j = \sin\left(\vartheta_j - \mu\right) = \sum_{q=1}^{\infty} \frac{a_q}{q!} H_q\left(Z_j\right)$$

are of the form

$$\gamma_Y(k) = \sum_{q=1}^{\infty} \frac{a_q^2}{q!} \gamma_Z^q(k) \underset{k\to\infty}{\sim} a_1^2 \gamma_Z(k)$$

and

$$f_Y(\lambda) \underset{\lambda\to 0}{\sim} c_{f,Y} |\lambda|^{-2d},$$

where $c_{f,Y} = a_1^2 c_{f,Z}$. Thus, asymptotically (10) is equivalent to

$$\hat{\mu} \pm \arcsin\left(n^{d-1/2}\sqrt{\nu(d)\, c_{f,Y}} z_{1-\alpha/2}\right). \tag{12}$$

Consistent estimation of $c_{f,Y}$ and d from an observed series $Y_1, \ldots, Y_n$ is well developed in the literature. The best-known methods include the Geweke Porter–Hudak estimator (Geweke and Porter-Hudak, 1983; Robinson, 1995a, 1995b; Moulines and Soulier, 1999), the local Whittle estimator (Künsch, 1987; Robinson, 1995b), the log-wavelet estimator (see e.g. Abry *et al.*, 2003; Faÿ *et al.*, 2009 and references therein), broadband FEXP estimation (Moulines and Soulier, 1999, 2000; Hurvich, 2001; Hurvich and Brodsky, 2001; Hurvich *et al.*, 2002;

 © 2019 John Wiley & Sons Ltd *J. Time Ser. Anal.* **41**: 210–228 (2020)
DOI: 10.1111/jtsa.12500

Narukawa and Matsuda, 2011; also see Beran, 1993; Robinson, 1994). In our context, μ is unknown so that estimation of $c_{f,Y}$ and d has to be based on $\hat{Y}_j = \sin\left(\vartheta_j - \hat{\mu}\right)$ $(j = 1, \ldots, n)$ instead of $Y_j = \sin(\vartheta_j - \mu)$. Since $\hat{\mu}$ is a consistent estimator, replacing Y_j by $\hat{Y}_j$ does not affect consistency of $\hat{c}_{f,Y}$ and $\hat{d}$. A data based confidence interval for μ can therefore be defined by

$$\hat{\mu} \pm \arcsin\left(n^{\hat{d}-1/2}\sqrt{v\left(\hat{d}\right)\hat{c}_{f,Y}}z_{1-\alpha/2}\right), \tag{13}$$

where $\hat{d}$ and $\hat{c}_{f,Y}$ are consistent estimators of d and $c_{f,Y}$ respectively.

3.3. Extension to Non-Stationary Series

The definition of the mean direction in Equation (1) implies that ϑ_j can be written as

$$\vartheta_j = \left(\mu + \varepsilon_j\right) \bmod 2\pi,$$

where $E(\sin \varepsilon_j) = 0$ and $E(\cos \varepsilon_j) = R$. More generally, we may assume a non-stationary model

$$\vartheta_j = \left[\mu\left(x_j; \beta\right) + \varepsilon_j\right] \bmod 2\pi, \tag{14}$$

where x_j $(j = 1, 2, \ldots, n)$ are explanatory variables observed at time j, β is an unknown parameter vector, ε_j $(j \in \mathbb{Z})$ is a stationary process such that

$$E\left(\sin \varepsilon_j\right) = 0, E\left(\cos \varepsilon_j\right) = R, \tag{15}$$

and the process $Y_j = \sin \varepsilon_j$ has an autocovariance function

$$\gamma_Y(k) = cov\left(Y_j, Y_{j+k}\right) \sim c_{\gamma,Y}|k|^{2d-1} \ \ (k \to \infty)$$

and a spectral density function

$$f_Y(\lambda) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\gamma_Y(k)\,e^{-ik\lambda} \sim c_{f,Y}\cdot|\lambda|^{-2d} \ \ (\lambda \to 0), \tag{16}$$

where $d \in (0, \frac{1}{2})$ and $0 < c_{\gamma,Y}, c_{f,Y} < \infty$. To obtain an estimator of β we note that $\hat{\mu}$ defined by (6) minimizes the risk function $Q_n(\mu) = \sum[1 - \cos(\vartheta_j - \mu)]$. Note that for unit vectors $u(\vartheta) = (\cos\vartheta, \sin\vartheta)^T$ and $u(\mu) = (\cos\mu, \sin\mu)^T$ $(\vartheta, \mu \in [0, 2\pi))$ we have

$$\begin{aligned}1 - \cos(\vartheta - \mu) &= 1 - (\cos\vartheta\cos\mu + \sin\vartheta\sin\mu) \\ &= 1 - u^T(\vartheta)\,u(\mu)\end{aligned}$$

which is minimal, if and only if the directions defined by ϑ and μ coincide. In the non-parametric literature for circular data, $1 - \cos(\vartheta - \mu)$ is therefore often used as a 'distance' between two angles ϑ and μ (see e.g. Hall *et al.*, 1987). Minimizing $Q_n(\mu)$ with respect to μ can be generalized to estimating β in (14) by defining

$$\hat{\beta} = \arg\min \tilde{Q}_n(b), \tag{17}$$

 © 2019 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12500

where

$$\tilde{Q}_n(b) = \sum_{j=1}^{n} \left[1 - \cos\left(\vartheta_j - \mu\left(x_j; b\right)\right)\right].$$

For an analogous definition in a non-parametric context see for example Di Marzio *et al.* (2012). More specifically consider for instance

$$\mu\left(x_j; \beta\right) = \left(x_j^T \beta\right) \bmod 2\pi = \left(\sum_{l=0}^{p} \beta_l x_{jl}\right) \bmod 2\pi, \tag{18}$$

where $\beta = (\beta_0, \dots, \beta_p)^T \in \mathbb{R}^{p+1}$ and $x_j = (x_{j0}, \dots, x_{jp})^T \in \mathbb{R}^{p+1}$ $(j = 1, \dots, n)$ are deterministic explanatory vectors (see e.g. Gould, 1969). Then minimizing

$$\tilde{Q}_n(b) = \sum_{j=1}^{n} \left[1 - \cos\left(\vartheta_j - x_j^T b\right)\right], \tag{19}$$

leads to $p + 1$ equations

$$\Psi_n\left(\hat{\beta}\right) = \left[\Psi_{n,0}\left(\hat{\beta}\right), \dots, \Psi_{n,p}\left(\hat{\beta}\right)\right]^T = 0, \tag{20}$$

where

$$\Psi_{n,l}\left(\hat{\beta}\right) = \sum_{j=1}^{n} x_{jl} \sin\left(\vartheta_j - x_j^T \hat{\beta}\right), \quad l = 0, \dots, p.$$

By standard arguments based on a Taylor expansion, $\hat{\beta} - \beta$ can be approximated by

$$\hat{\beta} - \beta = -\left\{E\left[\dot{\Psi}_n(\beta)\right]\right\}^{-1} \Psi_n(\beta) + o_p\left(\hat{\beta} - \beta\right), \tag{21}$$

where

$$\dot{\Psi}_n(\beta) = \left[\dot{\Psi}_{n;lm}(\beta)\right]_{l,m=0,\dots,p}$$

and

$$\dot{\Psi}_{n;lm}(\beta) = -\sum_{j=1}^{n} x_{jl} x_{jm} \cos\left(\varepsilon_j\right).$$

A detailed derivation and further simplifications can be obtained under suitable conditions on the explanatory variables. Denote by $X_n = [x_{jl}]_{j=1,\dots,n;l=0,\dots,p}$ the matrix of explanatory variables, with row vectors $x_j^T \in \mathbb{R}^{p+1}$ $(j = 1, \dots, n)$ and column vectors $x_{\cdot l}(n) = (x_{1l}, \dots, x_{nl})^T \in \mathbb{R}^n$ $(l = 0, \dots, p)$. Furthermore let $\|x_{\cdot l}(n)\| = \sqrt{\sum_{j=1}^{n} x_{jl}^2}$ and $D_n = diag(\|x_{\cdot 0}(n)\|, \dots, \|x_{\cdot p}(n)\|)$. The following conditions introduced by Grenander and Rosenblatt (1957) in the context of linear time series regression are useful:

- (R1) $\lim_{n\to\infty} \|x_{\cdot l}(n)\|^2 = \infty$ $(l = 0, \dots, p)$.
- (R2) $\lim_{n\to\infty} \|x_{\cdot l}(n)\|^2 / \|x_{\cdot l}(n-1)\|^2 = 1$ $(l = 0, \dots, p)$.

 © 2019 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12500

- (R3) The limits

$$\lambda_{lm}(k) = \lim_{n\to\infty} \frac{\sum_{j=1}^{n} x_{j,l}x_{j+k,m}}{\|x_{\cdot l}(n)\| \|x_{\cdot m}(n)\|} \quad (l, m = 0, \ldots, p;\ k \in \mathbb{N})$$

exist and are finite.
- (R4) Define $\lambda_{lm}(-k) = \lambda_{lm}(k)$ for $k \in \mathbb{N}$, and $(p+1)\times(p+1)$ matrices $\Lambda(k) = [\lambda_{lm}(k)]_{l,m=0,\ldots,p}$ $(k \in \mathbb{N})$. Then $\Lambda(0)$ is non-singular.

Assumptions (R1) to (R4) imply the following approximation for the variance matrix of $\hat{\beta} - \beta$:

Proposition 1. Let

$$V_n = \left[V_{n;l,m}\right]_{l,m=1,\ldots,p} \tag{22}$$

with

$$V_{n;l,m} = \sum_{j_1,j_2=1}^{n} \frac{x_{j_1 l}}{\left\|x_{\cdot l}(n)\right\|} \frac{x_{j_2 m}}{\left\|x_{\cdot m}(n)\right\|} \gamma_Y\left(j_1 - j_2\right). \tag{23}$$

Then, under (R1) to (R4),

$$D_n var\left(\hat{\beta}\right) D_n = R^{-2}\Lambda^{-1}(0)\, V_n \Lambda^{-1}(0) + r_n, \tag{24}$$

where r_n is of a smaller order than $D_n var(\hat{\beta}) D_n$.

From (24) in Proposition 1 one can see that the asymptotic variance of $\hat{\beta}$ depends on the autocovariance function, or equivalently the spectral density, of the process $Y_j = \sin \varepsilon_j$. Moreover, the asymptotic behaviour of V_n also depends on the so-called regression spectrum of the explanatory variables x_{jl} (see Yajima, 1991). For a detailed study of V_n under assumption (16) see Yajima (1988, 1991). To illustrate which results are possible, consider for instance a polynomial and a seasonal trend function respectively. For a polynomial trend

$$\mu_{\text{poly}}\left(x_j; \beta\right) = \left(\beta_0 + \sum_{l=0}^{p} \beta_l j^l\right) \bmod 2\pi$$

we have $D_n^2 \approx n \cdot diag[1, n/2, \ldots, n^p/(p+1)]$ and $\Lambda(0) = [\lambda_{l,m}(0)]_{l,m=0,\ldots,p}$ with

$$\lambda_{l,m}(0) = \frac{\sqrt{(2l+1)(2m+1)}}{l+m+1} \quad (j, l = 0, \ldots, p).$$

Yajima (1988) showed that in this case

$$\lim_{n\to\infty} n^{-2d} V_n = V = \left[v_{l,m}\right]_{l,m=0,\ldots,p},$$

where

$$v_{l,m} = 2\pi c_{f,Y} \frac{\sqrt{(2l+1)(2m+1)}\Gamma\left(\frac{3}{2} - H_Z\right)}{\Gamma\left(H_Z - \frac{1}{2}\right)\Gamma\left(\frac{3}{2} - H_Z\right)} \int_0^1 \int_0^1 u^l v^m |u - v|^{2H_Z - 2}\, \mathrm{d}u\mathrm{d}v.$$

© 2019 John Wiley & Sons Ltd

DOI: 10.1111/jtsa.12500

Thus

$$\lim_{n\to\infty} n^{-2d} D_n var\left(\hat{\beta}\right) D_n = R^{-2}\Lambda^{-1}(0)\, V\Lambda^{-1}(0). \tag{25}$$

Note in particular that the rate of convergence depends on the long-memory parameter d. Also, due to the linearization in (21), asymptotic normality of $n^{-d}D_n^{1/2}(\hat{\beta}-\beta)$ can be derived along the same line as for linear regression models (for detailed technical conditions see e.g. Yajima, 1991, and chapter 11 in Giraitis *et al.*, 2012 and references therein).

A very different result holds for a seasonal trend function defined by

$$\mu_{\text{season}}\left(x_j;\beta\right) = \left[\sum_{l=1}^{q}\beta_l \sin\left(\lambda_l j\right) + \sum_{l=1}^{q}\beta_{q+l}\cos\left(\lambda_l j\right)\right] \bmod 2\pi,$$

where $T \in \{2, 3, \dots\}$ and $\lambda_l = 2\pi l/T$ $(l = 1, \dots, q)$. Here, $\Lambda(0) = diag(1, \dots, 1)$, $D_n^2 \approx n \cdot diag(1/2, \dots, 1/2)$, and results in Yajima (1991) imply

$$\lim_{n\to\infty} V_n = V$$

with

$$V = 2\pi \cdot diag\left[f_Y\left(\lambda_1\right), \dots, f_Y\left(\lambda_q\right), f_Y\left(\lambda_1\right), \dots, f_Y\left(\lambda_q\right)\right].$$

Thus,

$$\lim_{n\to\infty} n \cdot var\left(\hat{\beta}-\beta\right) = 4\pi \cdot diag\left[f_Y\left(\lambda_1\right), \dots, f_Y\left(\lambda_q\right), f_Y\left(\lambda_1\right), \dots, f_Y\left(\lambda_q\right)\right].$$

Moreover, due to (21), under mild moment conditions $\sqrt{n}(\hat{\beta}-\beta)$ is asymptotically normal (see e.g. Yajima, 1991; Giraitis *et al.*, 2012). Note in particular that, in contrast to a polynomial trend, the rate of convergence of $\hat{\beta}$ is the same as under independence, for all values of the long-memory parameter d.

More generally, we may combine polynomial and seasonal trend functions,

$$\begin{aligned}\mu\left(x_j;\beta\right) &= \mu_{\text{poly}}\left(x_j;\beta_{\text{poly}}\right) + \mu_{\text{season}}\left(x_j;\beta_{\text{season}}\right)\\ &= \left[\sum_{l=0}^{p}\beta_{\text{poly},l} j^l + \sum_{l=1}^{q}\beta_{\text{season},l}\sin\left(\lambda_l j\right) + \sum_{l=1}^{q}\beta_{\text{season},q+l}\cos\left(\lambda_l j\right)\right] \bmod 2\pi.\end{aligned}$$

Analogous arguments as in Yajima (1991) imply that the estimated coefficients $\hat{\beta}_{\text{poly},0}, \dots, \hat{\beta}_{\text{poly},p}$ are asymptotically independent from $\hat{\beta}_{\text{season},1}, \dots, \hat{\beta}_{\text{season},2q}$.

With respect to confidence intervals for β and the function $\mu(x;\beta)$, the situation is more complicated than in the stationary case. The reason is that, if μ is not constant, then no explicit solution of (20) is available. Therefore (10) is no longer applicable. Instead, a linearization as indicated in (21) has to be used. For instance for $\mu = \mu_{\text{season}}$, $\sqrt{n}(\hat{\beta}-\beta) = \sqrt{n}(\hat{\beta}_1-\beta_1, \dots, \hat{\beta}_{2q}-\beta_{2q})$ is approximately distributed like a zero mean normal vector $Z = (Z_1, \dots, Z_{2q})$ with independent components and $var(Z_l) = var(Z_{l+q}) = 4\pi f_Y\left(\lambda_l\right)$ $(l = 1, \dots, q)$. A $(1-\alpha)$–confidence interval for individual coefficients β_l or β_{l+q} $(1 \le l \le q)$ is then given by $\hat{\beta}_l \pm q(\alpha)\sqrt{4\pi f_Y\left(\lambda_l\right)}$ and $\hat{\beta}_{l+q} \pm q(\alpha)\sqrt{4\pi f_Y\left(\lambda_l\right)}$ respectively, where $q(\alpha)$ is the $(1-\alpha/2)$-quantile of the standard normal distribution. In practice, $f_Y\left(\lambda_l\right)$ has to replaced by an estimate $\hat{f}_Y\left(\lambda_l\right)$ which can be obtained, as described above, by fitting a suitable model to the process $\hat{Y}_j = \sin\hat{\varepsilon}_j = \sin(\vartheta_j - \mu(x_j;\hat{\beta}))$. Moreover, noting that $\tilde{Z}_l = Z_l/\sqrt{4\pi f_Y\left(\lambda_l\right)}$ $(l = 1, \dots, 2q)$

 © 2019 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12500

are i.i.d. N(0, 1) variables, a (simultaneous) confidence interval for the vector $\beta = (\beta_1, \dots, \beta_{2q})^T$ can be defined by

$$C_\beta(\alpha) = \left\{ \beta^* \in \mathbb{R}^{2q} : \left| \frac{\hat{\beta}_l - \beta_l^*}{\sqrt{4\pi \hat{f}_Y(\lambda_l)}} \right| \le q(\alpha^*), l = 1, \dots, 2q \right\}, \tag{26}$$

where α^* is such that $(1-\alpha^*)^{2q} = 1-\alpha$. A (simultaneous) confidence band for $\mu_{\text{season}} = (\mu_{\text{season}}(1), \dots, \mu_{\text{season}}(n))^T$ is then given by

$$C_\mu(\alpha) = \left\{ y^* \in [0, 2\pi)^n : y_j^* = y_j \bmod 2\pi,\ \hat{\mu}_{low}(x_j) \le y_j \le \hat{\mu}_{up}(x_j) \right\}, \tag{27}$$

where

$$\hat{\mu}_{low}(x_j) = \min_{\beta^* \in C_\beta(\alpha)} \mu(x_j, \beta^*),\ \hat{\mu}_{up}(x_j) = \max_{\beta^* \in C_\beta(\alpha)} \mu(x_j, \beta^*)\ (j = 1, \dots, n). \tag{28}$$

Remark 4. Analogous asymptotic results for $\hat{\beta}$ in (17) can be derived for other definitions of μ. For instance Fisher and Lee (1992) consider $\mu(x) = \beta_0 + 2\arctan(x^T\beta)$. Other definitions of $\mu(x;\beta)$ can be found for instance in Johnson and Wehrly (1978). Also see for example Mardia and Jupp (1999), Jammalamadaka and SenGupta (2001), Pewsey and Neuhauser (2013), Kim and SenGupta (2016), and references therein.

4. SIMULATIONS

To examine finite sample properties of confidence intervals defined by (9) (known nuisance parameters) and (13) (estimated nuisance parameters) respectively, we consider model (4) with G equal to a von Mises and a wrapped normal distribution respectively. Recall that the density function $g_{\text{vM}} = G'$ of a von Mises distribution is defined by

$$g_{\text{vM}}(x) = g_{\text{vM}}(x; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp(\kappa \cos(x - \mu)),$$

where $I_0(\kappa)$ is the modified Bessel function of order 0, and $\kappa > 0$ is a scale parameter. The density function of a wrapped normal distribution is given by

$$g_{\text{wn}}(x) = g_{\text{wn}}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{k=-\infty}^{\infty} \exp\left(-\frac{(x - \mu + 2\pi k)^2}{2\sigma^2} \right).$$

Here $\sigma^2 > 0$ is the variance of the unwrapped normal distribution. Sometimes the so-called concentration parameter $\rho = \exp(-\sigma^2/2) \in (0, 1)$ is used for the wrapped distribution, instead of σ^2. The following specifications are used in the simulation study:

- Model 1: G =von Mises distribution with $\mu = \pi/5$ and $\kappa = 4$;
- Model 2: G =von Mises distribution with $\mu = \pi/5$ and $\kappa = 2$;
- Model 3: G =wrapped normal distribution with $\mu = \pi/5$ and concentration parameter $\rho = 0.5$.

The Gaussian process Z_j is generated by a standardized fractional ARIMA$(0, d, 0)$ process with $d = 0.2$ and 0.1 respectively. For each parameter setting, 100 simulated series are generated and 95%-confidence intervals (12) and (13) are calculated. The nuisance parameters $c_{f,Y}$ and d are estimated using the BIC criterion comparing ARIMA$(p, d, 0)$ with orders $p = 0, 1, \dots, [\log n]$ (see e.g. chapter 5 in Beran *et al.*, 2013). Simulated coverage probabilities of (12) and (13) are given in Tables I and II.

 © 2019 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12500

Table I. Models 1 and 2: simulated coverage probabilities of 95%-confidence intervals defined by (12) and (13) respectively

	Model 1					
	$c_{f,Y}$ and d known			$c_{f,Y}$ and d estimated		
	$n = 100$	400	1000	$n = 100$	400	1000
$d = 0.2$	0.98	0.96	0.94	0.87	0.91	0.94
$d = 0.4$	0.91	0.97	0.95	0.79	0.92	0.94
	Model 2					
	$n = 100$	400	1000	$n = 100$	400	1000
$d = 0.2$	0.96	0.97	0.90	0.79	0.95	0.91
$d = 0.4$	0.92	0.96	0.97	0.67	0.86	0.96

Table II. Model 3: simulated coverage probabilities of 95%-confidence intervals defined by (12) and (13) respectively

	$c_{f,Y}$ and d known				$c_{f,Y}$ and d estimated			
	$n = 100$	400	1000	4000	$n = 100$	400	1000	4000
$d = 0.2$	0.98	0.96	0.93	0.93	0.81	0.83	0.86	0.91
$d = 0.4$	0.85	0.95	0.90	0.95	0.68	0.84	0.83	0.90

Table III. Model 4: simulated coverage probabilities of 95%-confidence intervals defined by (13)

	$n = 100$	400	1000	4000
$d = 0.2$	0.85	0.90	0.95	0.92
$d = 0.4$	0.77	0.86	0.94	0.95

The results for Models 1 and 2 (Table I) illustrate that coverage probabilities of (12) are close to the nominal value of 0.95 for all sample sizes. On the other hand, estimation of $c_{f,Y}$ and d introduces additional uncertainty that is not negligible, if the series is short. The finite sample effect is more pronounced for $d = 0.4$. For larger sample sizes ($n = 1000$), (13) provides a reasonable approximation.

Similar comments apply to the wrapped normal distribution (Table II). In comparison with the von Mises distribution, estimation of $c_{f,Y}$ and d appears to have a stronger finite sample effect. Even for $n = 1000$, coverage probabilities of (13) are too low. We therefore carried out additional simulations for $n = 4000$. For this sample size, coverage probabilities are reasonably close.

Models 1–3 have symmetric unimodal marginal density functions. For comparison it is also interesting to consider a skewed bimodal density function, even though a mean direction in the sense of (1) may be less meaningful when there are two modes. Specifically we simulate confidence intervals with estimated nuisance parameters for (4) with

$$g(x) = G'(x) = \frac{4}{5} g_{\mathrm{vM}}\left(x; \mu_1, \kappa_1\right) + \frac{1}{5} g_{\mathrm{vM}}\left(x; \mu_2, \kappa_2\right), \tag{29}$$

where $\mu_1 = \pi/2$, $\mu_2 = \pi$, $\kappa_1 = 6$ and $\kappa_2 = 4$. In this case, the mean direction as defined by (1) is equal to $\mu = 1.8$. The simulation results in Table III (Model 4) show a reasonably good approximation of the nominal coverage probability for $n \geq 1000$. An interesting open question that should be addressed in future is how to estimate the more meaningful parameters μ_1 and μ_2.

5. DATA EXAMPLE

We consider daily average wind directions recorded in Milwaukee between 30 January and 31 December, 2017 (NOAA/GLERL Milwaukee WI Met Station, 43° 02′44″N, 87° 52′44″W; data source://www.glerl.noaa.gov/).

 © 2019 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12500

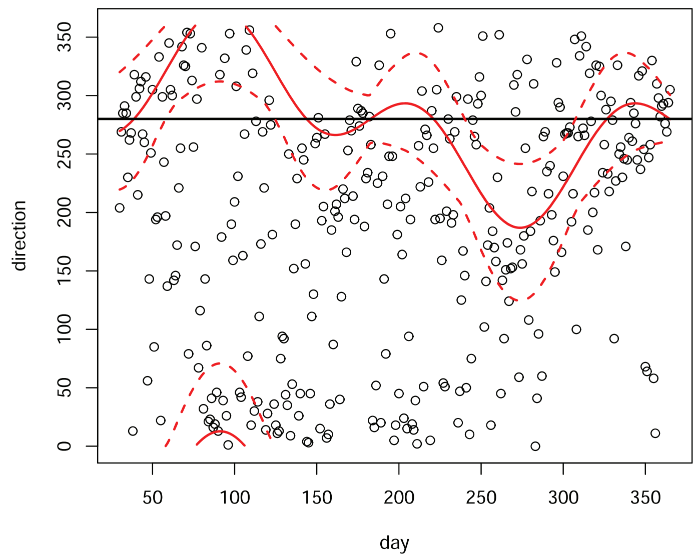

Figure 4. Milwaukee daily wind directions: time series plot of ϑ_j together with the fitted seasonal mean direction [Color figure can be viewed at wileyonlinelibrary.com]

Figure 1 shows 24 plots of the data, each corresponding to a 2-week period with 14 daily average wind directions. Note that in this data set, angles are measured clockwise and winds from north to south are represented by $\vartheta = 0°$. The time series plot of $\hat{Y}_j = \sin(\vartheta_j - \hat{\mu})$ is shown in Figure 2(a), the corresponding empirical autocorrelation function is given in Figure 2(c). Fitting a FARIMA$(p, d, 0)$ model using the BIC with a maximal order of $p_{\max} = [\log n]$ yields a FARIMA$(0, d, 0)$ process with $\hat{d} = 0.14$ and 95%-confidence $[0.06, 0.22]$. Thus there is evidence for long memory in the series. The periodogram of $\hat{Y}_j$ (in log–log-coordinates) together with the fitted spectral density is plotted in Figure 2(d). A windrose plot of the entire series is displayed in Figure 2(b). The estimated average direction is $\hat{\mu} = 280°$ which essentially corresponds to west winds (west to east). Based on the fitted model and (13), an approximate 95%-confidence interval for μ is $[233.2°, 326.8°]$. The confidence region is marked as a shaded area in Figure 2(b). Note that this confidence interval is much larger than the interval $[275.6°, 284.4°]$ obtained from formula (9) under the assumption of independence. In view of long memory observed in the data, the coverage probabilty of the shorter interval is however likely to be much lower than 0.95.

Given the geographic location of Milwaukee, the result can be interpreted as an overall dominance of a land breeze. The confidence interval for μ is however obtained under the assumption of stationarity. Often wind direction depends on the season. We therefore consider a seasonal model with period $T = 365$, and $\mu = [\beta_0 + \sum_{l=1}^{p} \beta_l \sin(\lambda_l j) + \sum_{l=1}^{p} \beta_{p+l} \cos(\lambda_l j)] \bmod 2\pi$ for a suitable value of $p \geq 1$. Fitting μ for $1 \leq p \leq 18$, and excluding non-significant coefficients (at the 5% level) leads to the parsimonious model $\hat{\mu}(j) = [\hat{\beta}_0 + \hat{\beta}_1 \sin(\lambda_1 j) + \hat{\beta}_3 \sin(\lambda_3 j)] \bmod 2\pi$. Note also that including an additional polynomial does not change the result, that is a polynomial trend is not significantly different from zero and can therefore be omitted. For $\hat{Y}_j = \sin(\vartheta_j - \hat{\mu}(x_j; \hat{\beta}))$, a FARIMA fit (as described above) yields $\hat{d} = 0.08$ with a 95%-confidence interval $[0, 0.16]$. Thus including a seasonal component in the model slightly reduces long memory in the stochastic component. The estimated coefficients for μ and the corresponding 95%-confidence intervals are $\hat{\beta}_0 = 279.8$ ($[259.1, 300.5]$), $\hat{\beta}_1 = 55.4$ ($[29.4, 81.4]$) and $\hat{\beta}_3 = -37.5$ ($[-61.3, -13.7]$). Note in particular that $\hat{\beta}_0 = 279.8$ is almost equal to $\hat{\mu} = 280°$ obtained under the assumption of stationarity, but the confidence interval for β_0 is shorter. The reason is that $\hat{\beta}_1 \sin(\lambda_1 j)$ and $\hat{\beta}_3 \sin(\lambda_3 j)$ capture seasonal deviations from the overall mean. The remaining variability is therefore smaller than before. Figure 4 shows ϑ_j together with the fitted seasonal trend. The full horizontal line corresponds to $\hat{\beta}_0 = 279.8$. Also given are simultaneous 95%-confidence bands for $\mu(j, \beta)$. The lower and upper border of the

© 2019 John Wiley & Sons Ltd

DOI: 10.1111/jtsa.12500

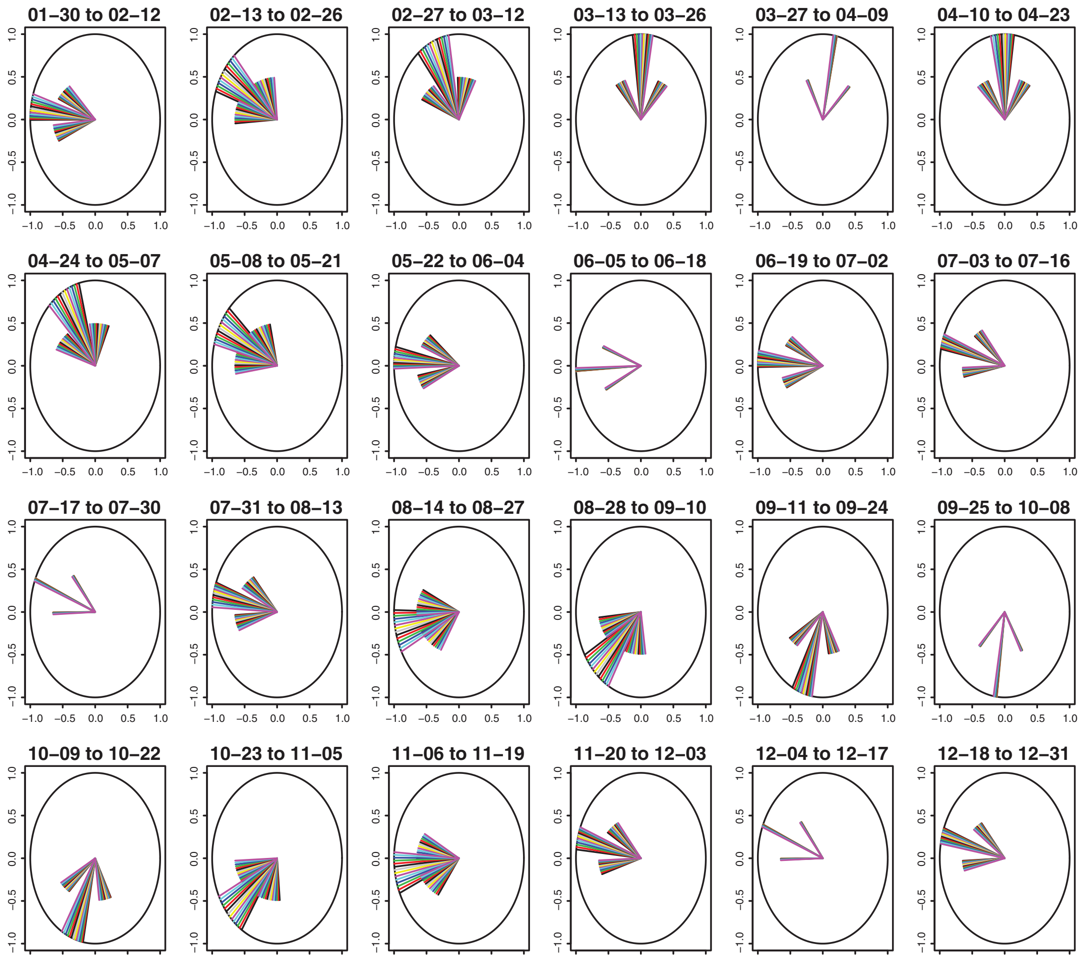

Figure 5. Seasonal trend $\hat{\mu}(j) = \hat{\beta}_0 + \hat{\beta}_1 \sin(\lambda_1 j) + \hat{\beta}_3 \sin(\lambda_3 j)$ of Milwaukee daily average wind directions between 30 January and 31 December, 2017. Each plot corresponds to a 2-week period. For each day, the wind direction defined by $\hat{\mu}(j)$ is displayed by an arrow. Thus, in each plot there are 14 arrows [Color figure can be viewed at wileyonlinelibrary.com]

confidence band are plotted as dotted red lines. Finally, in Figure 5, the daily mean wind directions defined by $\mu(j, \hat{\beta})$ are displayed for 24 consecutive 2-week periods. For each 2-week period, 14 arrows are drawn. Each arrow corresponds to the direction defined by $\mu(\cdot, \hat{\beta})$ on a particular day. The figures indicate a seasonal pattern with a dominance of north/north-west winds in spring, west winds in summer, south/south-west winds in fall, and west winds in winter. This pattern is close to long term weather patterns observed in Milwaukee.

6. FINAL REMARKS

In this article, Gaussian subordination was used to define a class of directional processes that exhibit long-range dependence. Given any continuous circular distribution function, a strongly dependent subordination model defined by (4) exists. More generally, one may define processes subordinated to linear processes that are not necessarily Gaussian. Note however that in the non-Gaussian case this leads to Appell polynomial expansions that are no longer orthogonal.

 © 2019 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12500

A further generalization one may consider is an extension to higher Hermite ranks. One of the assumptions in (12) and (13) is that $H(z) = \sin\{G^{-1}(\Phi(z)) - \mu\}$ has Hermite rank $m = 1$. For $m \geq 2$, the convergence of sums $\sum H(Z_j)$ is well understood so that analogous confidence intervals could be defined in principle. However, the limiting distribution is no longer normal, and quantiles are quite difficult to compute. In many applications $m = 1$ is a reasonable assumption. For example any monotonous transformation has Hermite rank one (see e.g. Menendez *et al.*, 2013). Moreover, tests of the null hypothesis that m is one are available in the literature (Beran *et al.*, 2016; Tewes, 2018).

The method proposed here enables us to model stationary and non-stationary circular time series data, including the estimation of trends and seasonal patterns, under general conditions on the temporal dependence structure. For instance, in the example discussed above, simultaneous confidence intervals for the seasonal wind direction were obtained. Modelling and prediction of the prevailing wind direction, together with an assessment of the uncertainty of estimates, are essential for many purposes, including weather forecast, climatology, wind power generation or air traffic. An interesting extension that will be worth pursuing in future research is the possibility of heteroskedasticity. For instance, the variability of wind direction may depend on the season or may even change at some point due to changes in the landscape. Another interesting question is how to carry out inference for a vector of mean directions in the case of mixture distributions.

ACKNOWLEDGEMENTS

We thank the referees for their insightful and constructive comments. The data set was obtained from the homepage of the NOAA Great Lakes Environmental Research Laboratory (https://www.glerl.noaa.gov/).

DATA AVAILABILITY STATEMENT

The data set was obtained from NOOA GLERL (NOAA – Great Lakes Environmental Research Laboratory; https://www.glerl.noaa.gov/). It is also available via the Supporting Information tab for this article.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

REFERENCES

Abry P, Flandrin P, Taqqu MS, Veitch D. 2003. Self-similarity and long-range dependence through the wavelet lens. In *Long-range Dependence: Theory and Applications*, Doukhan P., Oppenheim G., Taqqu M.S. (eds.) Boston: Birkhäuser; 527–556.

Beran J. 1993. Fitting long-memory models by generalized linear regression. *Biometrika* **80**: 817–822.

Beran J. 1994. *Statistics for Long-Memory Processes. Monographs on Statistics and Applied Probability*, Vol. 61. New York: Chapman and Hall/CRC.

Beran J, Feng Y, Ghosh S, Kulik R. 2013. *Long Memory Processes: Probabilistic Properties and Statistical Methods*. Berlin/Heidelberg: Springer.

Beran J, Möhrle S, Ghosh S. 2016. Testing for Hermite rank in Gaussian subordination processes. *Journal of Computational and Graphical Statistics* **25**(3): 917–934.

Breckling J. 1989. *The Analysis of Directional Time Series: Application to Wind Speed and Direction*. Berlin: Springer.

Di Marzio M, Panzera A, Taylor CC. 2012. Non-parametric smoothing and prediction for nonlinear circular time series. *Journal of Time Series Analysis* **33**(4): 620–630.

Dobrushin RL. 1980. Gaussian random fields—Gibbsian point of view. In *Multicomponent Random Systems. Advances in Probability and Related Topics*, Vol. 6, Dobrushin RL., Sinai YG. (eds.). New York: Dekker; 119–152.

Dobrushin RL, Major P. 1979. Non-central limit theorems for nonlinear functionals of Gaussian fields. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **50**(1): 27–52.

Faÿ G, Moulines E, Roueff F, Taqqu MS. 2009. Asymptotic normality of wavelet estimators of the memory parameter for linear processes. *Journal of Time Series Analysis* **30**(5): 534–558.

Fisher NI, Lee A. 1992. Regression models for an angular response. *Biometrics* **48**(3): 665–677.

 © 2019 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12500

Fisher NI. 1993. *Statistical Analysis of Circular Data*. Cambridge: Cambridge University Press.

Fisher NI, Lee AJ. 1994. Time series analysis of circular data. *Journal of the Royal Statistical Society. Series B* **56**(2): 327–339.

Fisher NI, Lewis T. 1983. Estimating the common mean direction of several circular or spherical distributions with differing dispersions. *Biometrika* **70**(2): 333–341.

Geweke J, Porter-Hudak S. 1983. The estimation and application of long memory time series models. *Journal of Time Series Analysis* **4**: 221–238.

Giraitis L, Koul HL, Surgailis D. 2012. *Large sample inference for long memory processes*. London: Imperial College Press.

Gould AL. 1969. A regression technique for angular variates. *Biometrics* **25**: 683–700.

Grenander CWJ, Rosenblatt M. 1957. *Statistical Analysis of Stationary Time Series*. New York: Academic Press.

Hall P, Watson GS, Cabrera J. 1987. Kernel density estimation with spherical data. *Biometrika* **74**(4): 751–762.

Hurvich CM. 2001. Model selection for broadband semiparametric estimation of long memory in time series. *Journal of Time Series Analysis* **22**: 679–709.

Hurvich CM, Brodsky J. 2001. Broadband semiparametric estimation of the memory parameter of a long-memory time series using fractional exponential models. *Journal of Time Series Analysis* **22**: 221–249.

Hurvich CM, Moulines E, Soulier P. 2002. The FEXP estimator for potentially nonstationary linear time series. *Stochastic Processes and their Applications* **97**: 307–340.

Jammalamadaka SR, Sarma YR. 1988. A correlation coefficient for angular variables. In *Statistical Theory and Data Analysis*, Matusita K. (ed.) New York: North Holland; 349–364.

Jammalamadaka SR, SenGupta A. 2001. *Topics in Circular Statistics*. Singapore: World Scientific.

Johnson RA, Wehrly TE. 1978. Some angular-linear distributions and related regression models. *Journal of the American Statistical Association* **73**(363): 602–606.

Kato S. 2010. A Markov process for circular data. *Journal of the Royal Statistical Society, Series B* **72**: 655–672.

Kim S, SenGupta A. 2016. Multivariate-multiple circular regression. *Journal of Statistical Computation and Simulation* **87**: 1–15.

Künsch H.R. 1987. Statistical aspects of self-similar processes. In *Proc. 1st World Congress of the Bernoulli Society*, Vol. Vol. 1. Utrecht: VNU Science Press; 67—74.

Ley C, Verdebout T. 2017. *Modern Directional Statistics*. Boca Raton: Chapman & Hall/CRC.

Major PJ. 1981. *Lecture Notes in Mathematics: Vol. 849. Multiple Wiener–Itô Integrals*. New York: Springer.

Mardia KV. 1972. *Statistics of Directional Data*. London: Academic Press.

Mardia KV, Jupp PE. 1999. *Directional Statistics*. New York: Wiley.

Menendez P, Ghosh S, Künsch HR, Tinner W. 2013. On trend estimation under monotone Gaussian subordination with long-memory: application to fossil pollen series. *Journal of Nonparametric Statistics* **25**(4): 765–785.

Modlin D, Fuentes M, Reich B. 2012. Circular conditional autoregressive modeling of vector fields. *Environmetrics* **23**: 46–53.

Moulines E, Soulier P. 1999. Broadband log-periodogram regression of time series with long-range dependence. *The Annals of Statistics* **27**: 1415–1439.

Moulines E, Soulier P. 2000. Data driven order selection for projection estimation of the spectral density of time series with long range dependence. *Journal of Time Series Analysis* **21**: 193–218.

Moulines E, Soulier P. 2003. Semiparametric spectral estimation for fractional processes. In *Theory and Applications of Long-Range Dependence*, Doukhan P., Oppenheim G., Taqqu MS. (eds.) Boston: Birkhäuser; 251–301.

Narukawa M, Matsuda Y. 2011. Broadband semi-parametric estimation of long-memory time series by fractional exponential models. *Journal of Time Series Analysis* **32**: 175–193.

Pewsey A, Neuhauser M. 2013. *Circular Statistics in R*. Oxford: Oxford University Press.

Robinson PM. 1994. Time series with strong dependence. In *Advances in Econometrics: Sixth World Congress*, Vol. 1, Sims CA (ed.). Cambridge: Cambridge University Press; 47–95.

Robinson PM. 1995a. Log-periodogram regression of time series with long range dependence. *The Annals of Statistics* **23**(3): 1048–1072.

Robinson PM. 1995b. Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics* **23**(5): 1630–1661.

Rosenblatt M. 1961. Independence and dependence. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob.*, University of California Press: Berkeley, 431–443.

Rosenblatt M. 1979. Some limit theorems for partial sums of quadratic forms in stationary Gaussian variables. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **49**(2): 125–132.

Taqqu MS. 1975. Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **31**: 287–302.

Taqqu MS. 1979. Convergence of integrated processes of arbitrary Hermite rank. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **50**: 53–83.

Tewes J. 2018. Block bootstrap for the empirical process of long-range dependent data. *Journal of Time Series Analysis* **39**(1): 28–53.

 © 2019 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12500

Wang F, Gelfand AE. 2014. Space and space-time directional data using projected Gaussian processes. *Journal of the American Statistical Association* **109**(508): 1565–1580.
Wehrly TE, Johnson RA. 1980. Bivariate models for dependence of angular observations and a related Markov process. *Biometrika* **66**: 255–256.
Yajima Y. 1988. On estimation of a regression model with long term errors. *The Annals of Statistics* **16**(2): 791–807.
Yajima Y. 1991. Asymptotic properties of the LSE in a regression model with long-memory stationary errors. *The Annals of Statistics* **19**: 158–177.

APPENDIX A. PROOFS

Proof of Lemma 1. Consider the space $L^2(\mathbb{R},\varphi)$ of real valued functions $H:\mathbb{R}\to\mathbb{R}$ with $\int H^2(z)\varphi(z)\mathrm{d}z<\infty$ where $\varphi(z)=(2\pi)^{-\frac{1}{2}}\exp\left(-\frac{1}{2}z^2\right)$. For functions $H,\tilde{H}\in L^2(\mathbb{R},\varphi)$, define the scalar product

$$\left\langle H,\tilde{H}\right\rangle=\int H(z)\,\tilde{H}(z)\,\varphi(z)\,\mathrm{d}z.$$

Then $L^2(\mathbb{R},\varphi)$ is a Hilbert space where Hermite polynomials $H_q(z)=(-1)^q\exp(\frac{1}{2}z^2)\,d^q/dz^q\exp(-\frac{1}{2}z^2)$ build an orthogonal basis. Any function $H\in L^2(\mathbb{R},\varphi)$ then has the L^2-representation

$$H(z)=\sum_{q=0}^{\infty}\frac{a_q}{q!}H_q(z)$$

with

$$a_q=\left\langle H_q,H\right\rangle=\int H_q(z)\,H(z)\,\varphi(z)\,\mathrm{d}z.$$

Moreover, H is called to be of Hermite rank $m\geq 1$, if $a_m\neq 0$ and $a_q=0$ $(q<m)$. Based on this representation, Taqqu (1975) showed that for a stationary zero mean Gaussian process Z_j with $var(Z_j)=1$, and autocovariances γ_Z and a spectral density f_Z that satisfy (3) with $d>\frac{1}{2}(1-m^{-1})$, one obtains

$$cov\left(H\left(Z_j\right),H\left(Z_{j+k}\right)\right)\underset{k\to\infty}{\sim}\frac{a_m^2}{m!}c_{\gamma,Z}^m k^{(2d-1)m} \tag{A1}$$

(also see e.g. Major, 1981). In particular, consider

$$H(z)=\sin\left\{G^{-1}(\Phi(z))-\mu\right\},$$

and suppose that H has Hermite rank m. Then (A1) holds with

$$a_m=\int H_m(z)\sin\left\{G^{-1}(\Phi(z))-\mu\right\}\varphi(z)\,\mathrm{d}z.$$

The result then follows by rewriting $(2d-1)m$ as $2d_m-1$. □

Proof of Theorem 1. Since

$$\bar{S}^*=n^{-1}\sum\sin\vartheta_j^*$$

and

$$H(z) = \sin\left\{G^{-1}(\Phi(z)) - \mu\right\}$$

has Hermite rank 1, the result follows from limit theorems under Gaussian subordination (Taqqu, 1975; Dobrushin and Major, 1979; also see e.g. Beran *et al.*, 2013, chapter 4.2). □

Proof of Proposition 1. Minimizing $\tilde{Q}_n$ leads to

$$\Psi_n\left(\hat{\beta}\right) = \left[\Psi_{n,0}\left(\hat{\beta}\right), \dots, \Psi_{n,p}\left(\hat{\beta}\right)\right]^T = 0$$

where

$$\begin{aligned}\Psi_n\left(\hat{\beta}\right) &= \left(\sum_{j=1}^{n} x_{j0}\sin\left(\vartheta_j - x_j^T\hat{\beta}\right), \dots, \sum_{j=1}^{n} x_{jp}\sin\left(\vartheta_j - x_j^T\hat{\beta}\right)\right)^T \\ &= X_n^T s_n\left(X_n, \hat{\beta}\right),\end{aligned}$$

X_n denotes the $n \times (p+1)$ matrix with elements x_{jl} ($j = 1, \dots, n$; $l = 0, \dots, p$) and

$$s_n\left(X_n, \hat{\beta}\right) = \left[\sin\left(\vartheta_1 - x_1^T\hat{\beta}\right), \dots, \sin\left(\vartheta_n - x_n^T\hat{\beta}\right)\right]^T.$$

Standard arguments based on a Taylor expansion lead to

$$\hat{\beta} - \beta = -\dot{\Psi}_n^{-1}(\beta)\,\Psi_n(\beta) + o_p\left(\hat{\beta} - \beta\right)$$

with

$$\dot{\Psi}_n(\beta) = \left[\dot{\Psi}_{n;lm}(\beta)\right]_{l,m=0,\dots,p}$$

and

$$\dot{\Psi}_{n;lm}(\beta) = -\sum_{j=1}^{n} x_{jl}x_{jm}\cos\left(\varepsilon_j\right).$$

Noting that, under (R1) to (R4), $D_n^{-1}\dot{\Psi}_n(\beta)D_n^{-1}$ can be approximated by $-R\Lambda(0)$, we obtain

$$\begin{aligned}D_n\left(\hat{\beta} - \beta\right) &= -\left[D_n^{-1}\dot{\Psi}_n(\beta)D_n^{-1}\right]^{-1} D_n^{-1}X_n^T s_n\left(X_n, \hat{\beta}\right) + o_p\left(\hat{\beta} - \beta\right) \\ &= R^{-1}\Lambda^{-1}(0)\, D_n^{-1}X_n^T s_n\left(X_n, \hat{\beta}\right) + o_p\left(\hat{\beta} - \beta\right).\end{aligned}$$

Hence,

$$D_n var\left(\hat{\beta}\right) D_n = R^{-2}\Lambda^{-1}(0)\, V_n\Lambda^{-1}(0) + o\left(D_n var\left(\hat{\beta}\right) D_n\right),$$

where V_n is as defined in (23). □

 © 2019 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12500