Journal of Econometrics 239 (2024) 105450

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Modelling circular time series☆

Andrew Harvey [a], Stan Hurn [d], Dario Palumbo [b,c,*], Stephen Thiele [d]

[a] Faculty of Economics, University of Cambridge, United Kingdom
[b] Department of Economics, Ca' Foscari University of Venice, Italy
[c] Homerton College, University of Cambridge, United Kingdom
[d] School of Economics and Finance, Queensland University of Technology, Australia

ARTICLE INFO

Article history:
Received 14 December 2021
Received in revised form 14 February 2023
Accepted 27 February 2023
Available online 29 May 2023

JEL classification:
C22

Keywords:
Directional statistics
Dynamic conditional score model
Nonstationarity
von Mises distribution
Wind direction

ABSTRACT

Circular variables often play an important role in the construction of models for analysing and forecasting the consequences of climate change and its impact on the environment. Such variables pose special problems for time series modelling. This article shows how the score-driven approach, developed primarily in econometrics, provides a natural solution to the difficulties and leads to a coherent and unified methodology for estimation, model selection and testing. The new methods are illustrated with data on wind direction.

1. Introduction

There is now a growing body of evidence that directly attributes natural disasters to climate change. This literature has been given fresh impetus by a major research initiative established in 2015. The World Weather Attribution project https://www.worldweatherattribution.org conducts real-time attribution analysis of extreme weather events, such as bushfires, floods, drought and heatwaves, as they happen around the world; see, for example, van Oldenborgh et al. (2021). As this evidence grows it becomes increasingly important that statistical methods developed by econometricians play a role in analysing and modelling the consequences and challenges of climate change.

Many scientific fields have applications in which circular data are collected and statistically analysed. This paper will be particularly concerned with wind direction. Not only is modelling and forecasting wind direction relevant for wind farms and for assessing the effects of pollution from a particular location, as in García-Portugués et al. (2013), but it is also a crucial input into tracking wildfires, the frequency of which is on the increase primarily because of climate change; see Clarke et al. (2013) and van Oldenborgh et al. (2021). Furthermore, recent articles, such as those by Lagona (2019) and Lobeto et al. (2022) have started to explore the climate-driven changes in wind-wave climatology. This area of research highlights not only the importance of modelling wind direction, but also admits analysis of another circular variable, wave direction.

☆ An early discussion of some of the ideas in this paper can be found in an unpublished discussion paper by Harvey et al. (2019).

* Corresponding author at: Department of Economics, Ca' Foscari University of Venice, Fondamenta S. Giobbe, Cannareggio 873, 30121 Venezia, Italy
E-mail addresses: ach34@cam.ac.uk (A. Harvey), dario.palumbo@unive.it (D. Palumbo).

https://doi.org/10.1016/j.jeconom.2023.02.016

1.1. Wind turbines and wildfires

Given the importance of wind power to the climate change agenda and given the critical importance of wind speed to electricity generated by the wind turbines, it is not surprising that there has been a good deal of academic literature focusing on predicting wind speed. What has been relatively neglected, however, is the role of wind direction, which is a crucial component of efficient turbine control; see Erdem and Shi (2011). For modern horizontal-axis turbines, wind direction is important because, as noted by Song et al. (2017), yaw[1] misalignment leads to decreased power capture that is proportional to the cube of the yaw error. Yaw errors can also lead to system failure. Conventional control methods are generally based on backward looking algorithms that look only at past errors and result in a control delay. Making use of predicted values may improve the performance of the yaw control system substantially.

Rapp et al. (2021) highlight the importance of short term forecasts of wind speed and direction for fire management, while Page et al. (2018) point out that American and Australian evidence suggests that the official forecasts need to be improved. As pointed out in Sharples et al. (2010), fire spread models are highly sensitive to changes in wind direction. Recent evidence from the United States suggests that fire managers use mental shortcuts that can bias decision-making. It is therefore imperative that crucial inputs into fire spread models, particularly, wind direction are the focus of modelling energy. While it may be tempting to relying on "official" wind forecasts from large scale models that are provided by government organizations, recent evidence from American and Australian on wind speed forecast suggest that this approach may be sub-optimal. In the U.S., Page et al. (2018) suggest that wild fire manages should exercise caution when using forecasts from the National Digital Forecast Database as these tend to underpredict wind speeds. In Australia, Hurn et al. (2021) document the variable performance of official Bureau of Meteorological forecasts of wind speed. While this evidence is cast in terms of wind speed there is no reason to expect any better performance in terms of wind direction.

1.2. Circular observations and time series models

Circular variables pose a challenge for time series modelling. If the starting point is due south then moving through 180 degrees ends up at due north. The same point is reached by moving 180 degrees in the opposite direction. In terms of radians the points $-\pi$ and π meet up. Hence observations slightly smaller than π and slightly bigger than $-\pi$ are very close whereas on a linear scale they are far apart. As a consequence, the arithmetic mean of points clustered around $\pm\pi$ will be close to zero, whereas the *mean direction* is close to $\pm\pi$. Mardia and Jupp (2000) provide a thorough discussion of the implications of circularity for statistical measures.

A number of ways for modelling circular time series have been proposed in the literature; see the recent review by Pewsey and García-Portugués (2021), section 10. Most of these approaches struggle to find a method that is tractable, straightforward to interpret and sufficiently general to handle a wide variety of situations. Forecasting is never discussed. Furthermore diagnostic checks are minimal with no paper reporting tests for residual serial correlation. The aim of the present paper is to develop a class of models that addresses these issues and yields a coherent model specification methodology that is straightforward to implement.

In an early influential paper, Fisher and Lee (1994) proposed a method based on transformations, the aim of which is to try to put the data in a form that lends itself to conventional autoregressive (AR) or autoregressive moving average (ARMA) modelling. These are known as linked ARMA (LARMA) models. The aim is to transform a circular variable in the range $[-\pi, \pi)$ to one on the real line that is close to being Gaussian. An ARMA model is then fitted in the usual way and the reverse transform is then applied to bring it back to the range of the original circular variable. Popular reverse transformations include probit and logistic. A second class of models, uses similar transformations on lagged observations to formulate an autoregressive model in which the conditional distribution of the observation is circular. These models are nonlinear and are called inverse autoregressions (IAR) or circular autoregressions (CAR). Both LARMA and IAR models have the limitation of being restricted to a narrow arc on the circle; see Fisher and Lee (1994), p 332. This being the case, it raises the question of why do the transformation in the first place.

As with IAR models, the models here are based on a circular conditional distribution, but the dynamics are formulated so as to be consistent with this distribution. The approach is closest in spirit to the one proposed by Breckling (1989) but it solves the computational problems that have limited its scope and applicability. The modelling draws on recently developed procedures for dealing with non-Gaussian conditional distributions in a wide variety of situations, primarily in economics and finance. The defining feature of the new class of circular time series models, which turns out to be crucial for performance as well as theoretical properties, is that the dynamics of the time-varying parameter are driven by the score of the conditional distribution. Score-driven models are known as Dynamic Conditional Score (DCS) or Generalized Autoregressive Score (GAS) models; see Harvey (2013) and Creal et al. (2013).

[1] The yaw system of wind turbines is the component responsible for the orientation of the wind turbine rotor towards the wind.

1.3. Plan of the paper

The plan of the paper is as follows. Section 2 reviews circular distributions, the most prominent of which is the von Mises distribution, and then goes on to show how score-driven models provide a natural solution to handling circular time series. The asymptotic distribution of the maximum likelihood (ML) estimator is derived for a first-order model and specification tests, based on the Lagrange multiplier principle, are proposed. An example highlights the shortcomings of ARMA and LARMA models when the observations are spread out over the circle.

Section 3 applies the ML estimation theory to a simple non-stationary model and shows how models based on differenced observations are formulated. The issues involved in choosing between models in levels or differences are investigated and illustrated with applications. Dynamic models for a general class of circular distributions are discussed in Section 4, with special emphasis on the wrapped Cauchy. The Section 5 concludes and points to future developments.

2. Score-driven models for circular data

Given circular data measured in radians over an interval of length 2π, a standard assumption is that the observations are from a von Mises (vM) distribution with probability density function (PDF) given by

$$f(y;\mu,\upsilon)=\frac{1}{2\pi I_0(\upsilon)}\exp\{\upsilon\cos(y-\mu)\},\qquad \upsilon\geq 0, \tag{1}$$

where $I_k(\upsilon)$ denotes a modified Bessel function of order k, μ denotes location (mean direction) and υ is a non-negative concentration parameter that is inversely related to dispersion. When $\upsilon=0$, the distribution is uniform and when υ is large, y is approximately normal with mean μ and variance $1/\upsilon$, that is $N(\mu,1/\upsilon)$. The normal distribution is generally considered a good approximation for $\upsilon>2$. The (circular) variance of the von Mises distribution is

$$1-I_1(\upsilon)/I_0(\upsilon)=1-A(\upsilon).$$

Note that $A(\upsilon)=E\cos(y-\mu)$ defines the mean resultant length and that $A(\upsilon)\to 1$ as $\upsilon\to\infty$, whereas $A(0)=0$.

More generally, following Mardia and Jupp (2000), p 26, a (continuous) circular PDF which depends on a vector of parameters $\boldsymbol{\theta}$, denoted $f(y;\boldsymbol{\theta})$, must satisfy the following conditions:

$$\text{(i) } f(y;\boldsymbol{\theta})\geq 0;\ \text{(ii) } \int_{-\pi}^{\pi} f(y;\boldsymbol{\theta})dy=1 \text{ and (iii) } f(y\pm 2\pi k;\boldsymbol{\theta})=f(y;\boldsymbol{\theta}),$$

where k is an integer and, for (i) and (iii), $-\infty<y<\infty$ almost everywhere; (iii) is the periodicity condition. A general class of circular distributions, which includes the vM as a special case, is proposed in Jones and Pewsey (2005), and investigated in Section 4. Ley and Verdebout (2017) and Pewsey and García-Portugués (2021), pp 4–7, discuss this and other classes of circular distributions.

2.1. Dynamic location

The observations generated by a Gaussian time series model over the real line, that is $-\infty<x_t<\infty$, can be converted into wrapped circular time series observations in the range $[-\pi,\pi)$ by letting

$$y_t=x_t \bmod(2\pi)-\pi,\qquad t=1,\ldots,T. \tag{2}$$

If the range $[0,2\pi)$ is preferred, there is no need to subtract π.

Fitting observations assumed to come from wrapped linear models can be seen as a missing data problem because x_t can be decomposed as

$$x_t=y_t+2\pi k_t,\quad t=1,\ldots,T, \tag{3}$$

where k_t is an integer that needs to be estimated; see Breckling (1989) and Fisher and Lee (1994), p 329. The difficulty with this approach is computational. Estimation is usually by the EM algorithm, but, as Fisher and Lee (1994), pp 333–4, observe, this can become complicated for all but the simplest models. Coles (1998) suggests using Markov chain Monte Carlo to fit autoregressive models.

The score-driven model is also set up for a variable defined over the real line. The idea can be introduced by first considering the basic unobserved components Gaussian model

$$\begin{aligned} x_t &= \mu+\mu_t+\epsilon_t,\quad t=1,\ldots,T,\\ \mu_t &= \phi\mu_{t-1}+\eta_t, \end{aligned} \tag{4}$$

where $-\infty<x_t<\infty$, μ is the mean, ϕ is the autoregressive parameter and η_t and ϵ_t are mutually independent normally distributed white noise disturbances with variances σ_η^2 and σ_ϵ^2 respectively. The steady-state innovations form of the

Kalman filter is

$$x_t = \mu + \mu_{t|t-1} + v_t, \qquad t = 1, \ldots, T, \tag{5}$$

$$\mu_{t+1|t} = \phi\mu_{t|t-1} + \kappa v_t,$$

where the second equation is the steady-state Kalman filter and κ is the Kalman gain. The innovations v_t are independently and identically distributed (IID) by construction. When $|\phi| < 1$, the model is stationary and $\mu_{t|t-1}$ is a linear combination of past observations with exponentially declining weights. If $\kappa = \phi$, the model reduces to a first-order autoregressive process, denoted $AR(1)$.

Score-driven non-Gaussian models take the observation-driven (5), rather than the parameter-driven (4), as the starting point. For circular data,

$$x_t = \mu + \mu_{t|t-1} + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{6}$$

where $-\pi \leq \mu < \pi$ and the $\varepsilon_t's$ are IID random variables each of which is assumed to have a circular distribution with location zero. The dynamic equation corresponding to the filter in (5) is

$$\mu_{t+1|t} = \phi\mu_{t|t-1} + \kappa u_t, \tag{7}$$

where the forcing variable, u_t, is defined as being (proportional to) the conditional score for location. The filter used in practice is of the same form but initialized with $\mu_{1|0}$ taking a fixed value, usually zero.

If x_t falls outside the range $[-\pi, \pi)$ it can be wrapped, as in (2), so that y_t lies in the range $[-\pi, \pi)$. The score-driven filter is invariant to this wrapping because the score is assumed to be continuous and so, like the conditional distribution, it satisfies the periodicity condition. Hence adding or subtracting multiples of 2π does not affect its value and so it can be expressed in terms of y_t instead of x_t. We therefore have the following proposition.

Proposition 1. *When x_t has a conditional circular distribution, the conditional distribution of y_t in a model defined by (6), (2) and (7) is the same as that of x_t.*

As a consequence of the above result, the likelihood function of the wrapped observations, the $y_t's$, is the same as that of the unobserved variables, the $x_t's$, and so the question of how to estimate the $k_t's$ in (3) does not arise. The proposition holds more generally when $\mu_{t|t-1}$ depends on any function of past scores with suitably defined starting values.

Remark 2. The filtered location, $\mu + \mu_{t|t-1}$, does not need to be wrapped, as in (2), but if it is, the conditional distribution of y_t is unaffected.

The scores in (7) are IID with mean zero and this property continues to hold when they are expressed as a function of $y_t - \mu_{t|t-1}$. In the case of the von Mises distribution, that is $\varepsilon_t \sim \text{vM}(0, \upsilon)$ in (6), the score is $\partial \ln f_t / \partial \mu_{t|t-1} = \upsilon \sin(y_t - \mu - \mu_{t|t-1})$. The concentration parameter, υ, may be dropped so the filter is no longer dependent on it. Thus

$$u_t = \sin(y_t - \mu - \mu_{t|t-1}). \tag{8}$$

These variables are $IID(0, A(\upsilon)/\upsilon)$ because $y_t - \mu - \mu_{t|t-1} = \varepsilon_t$.

For small deviations from the mean, u_t is approximately linear; the MacLaurin expansion shows that $\sin(y_t - \mu - \mu_{t|t-1}) \simeq y_t - \mu - \mu_{t|t-1}$. However, it starts to downweight observations for deviations beyond $\pm\pi/2$. When the concentration is large, so that a Gaussian conditional distribution is a good approximation, the model is close to the steady-state innovations form of the Kalman filter in (5).

The dots in Fig. 1 illustrate the critical role played by the score when $\mu = 0$ and $\mu_{t|t-1}$ is near π. Suppose $\mu_{t|t-1} = \pi - a$, where a is small and positive. (In the figure it is $a = \pi - 2$.) If the next observation is negative at $-\pi + b$, where b is small and positive, the distance between $\mu_{t|t-1}$ and y_t is only $a + b$, but $y_t - \mu_{t|t-1} = -2\pi + b + a$. However, $\sin(y_t - \mu_{t|t-1}) = \sin(-2\pi + b + a) = \sin(b + a)$ so the impact of the negative observation on $\mu_{t|t-1}$ is positive.

Forecasts of future values of y_t can be made recursively from the dynamic equation, as discussed in Harvey (2013), p 89. If a forecast falls outside the range $[-\pi, \pi)$ it can be wrapped. The conditional distribution for y_{T+1} is $vM(\widetilde{y}_{T+1|T}, \upsilon)$. The conditional distribution of $y_{T+\ell}$, $\ell \geq 2$, may be obtained by simulation.

Remark 3. The score-driven vM model may be extended to allow for heteroscedasticity by setting up a dynamic equation for $\upsilon_{t|t-1}$, the concentration parameter of ε_t in (6). The link function $\upsilon_{t|t-1} = \exp(\lambda_{t|t-1})$, $-\infty < \lambda_{t|t-1} < \infty$, ensures that $\upsilon_{t|t-1}$ remains positive. The implementation of this extension will be pursued elsewhere.

2.2. Maximum likelihood estimation

As was shown in Proposition 1, the likelihood function for a circular variable is unchanged when the $x_t's$ are replaced by the observable wrapped observations, y_t. Thus any asymptotic theory for x_t applies to estimates computed from observations in the range $[-\pi, \pi)$.

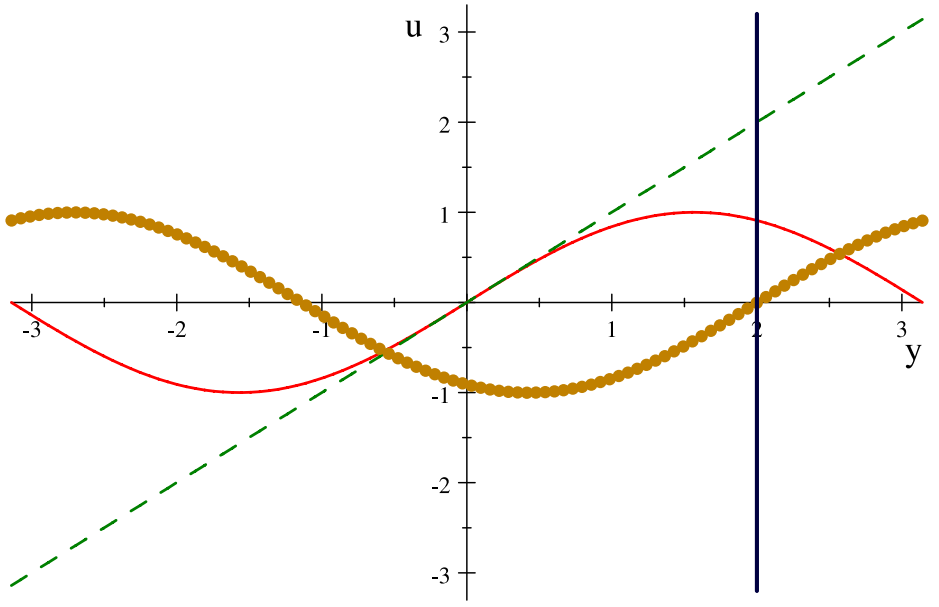

Fig. 1. Score functions for vM (solid line) and normal (dashed). Dots denote vM with $\mu = 2$.

Proposition 4. *For model (6) with $\varepsilon_t \sim vM(0, \upsilon)$, define $\boldsymbol{\psi} = (\kappa, \phi, \mu)'$ in the stationary first-order dynamic Eq. (7), with $|\phi| < 1$ and $\upsilon > 0$, and u_t as in (8). For a single observation, the information matrix for $\boldsymbol{\psi}$ and υ is*

$$\mathbf{I}\begin{pmatrix} \boldsymbol{\psi} \\ \upsilon \end{pmatrix} = \begin{bmatrix} \upsilon A(\upsilon)\mathbf{D}(\boldsymbol{\psi}) & \mathbf{0} \\ \mathbf{0}' & 1 - A(\upsilon)^2 - A(\upsilon)/\upsilon \end{bmatrix}, \tag{9}$$

where

$$\mathbf{D}(\boldsymbol{\psi}) = \mathbf{D}\begin{pmatrix} \kappa \\ \phi \\ \mu \end{pmatrix} = \frac{1}{1-b}\begin{bmatrix} A(\upsilon)/\upsilon & \dfrac{a\kappa A(\upsilon)/\upsilon}{1-a\phi} & 0 \\ \dfrac{a\kappa A(\upsilon)/\upsilon}{1-a\phi} & \dfrac{\kappa^2 A(\upsilon)(1+a\phi)/\upsilon}{(1-\phi^2)(1-a\phi)} & 0 \\ 0 & 0 & \dfrac{(1-\phi)^2(1+a)}{1-a} \end{bmatrix}$$

with $a = \phi - \kappa A(\upsilon)$ and

$$b = \phi^2 - 2\phi\kappa A(\upsilon) + \kappa^2(1 - A(\upsilon)/\upsilon) < 1. \tag{10}$$

Proof. The information matrix for a vM distribution with parameters μ and υ is given in Mardia and Jupp (2000), p 86 and 350. The general expression for $\mathbf{D}(\boldsymbol{\psi})$ is given in Harvey (2013), p 37. It is evaluated here by observing that $\sigma_u^2 = E(u_t^2) = A(\upsilon)/\upsilon$ and, with u_t' denoting $\partial u_t/\partial\mu$, $E\left(u_t'\right) = -E\cos\varepsilon_t = -A(\upsilon)$. Then, because the information quantity for υ is $E[(\cos\varepsilon_t - A(\upsilon))^2] = 1 - A(\upsilon)^2 - A(\upsilon)/\upsilon$, it follows that $E\left(u_t'\right)^2 = E\left(\cos^2\varepsilon_t\right) = 1 - A(\upsilon)/\upsilon$ and finally $c = -\kappa E\left(u_t u_t'\right) = -\kappa E(\sin\varepsilon_t\cos\varepsilon_t) = -\kappa E\left(\sin\{2\varepsilon_t\}\right)/2 = 0$. There are no extra terms because the off-diagonals in the information matrix for μ and υ are zero and u_t does not depend on υ.

Remark 5. A sufficient condition for strict stationarity and ergodicity of the derivatives of $\mu_{t|t-1}$ is $E\ln|z_t| < 0$, for all admissible $\boldsymbol{\psi}$, where $z_t = \phi + \kappa u_t'$; see Straumann and Mikosch (2006). By Jensen's inequality $E(\ln|z_t|) \leq \ln E(|z_t|)$ and hence $E(|z_t|) < 1$ is sufficient for $E(\ln|z_t|) < 0$ to hold. Unless z_t is positive, $E(|z_t|)$ can be difficult to evaluate. However, because the Cauchy–Schwarz inequality is

$$E(|z_t|^2) = E(z_t^2) = b \geq [E(|z_t|)]^2,$$

it follows that if $b < 1$ then $[E(|z_t|)]^2 < 1$ and so $E(|z_t|) < 1$.

Remark 6. When $\upsilon \to \infty$ the condition $b < 1$ leads to $(\phi - \kappa)^2 < 1$ which corresponds to the familiar condition for invertibility in an ARMA(1,1) model and the asymptotic distribution for estimators in a Gaussian model is obtained; see Harvey (2013), pp 67–8. On the other hand, as $\upsilon \to 0$, $b \to \phi^2 + \kappa^2/2$ because $A(\upsilon) \to 0$ and $A(\upsilon)/\upsilon \to 1/2$.

If the value of $\mu_{1|0}$ were given, Lemma 1 in Jensen and Rahbek (2004) could be invoked to show that when $\kappa \neq 0$, $0 < \upsilon < \infty$ and $b < 1$, the model is locally identifiable. In this instance the ML estimator, $(\widetilde{\boldsymbol{\psi}}' \ \widetilde{\upsilon})'$, is consistent, with the distribution of $\sqrt{T}(\widetilde{\boldsymbol{\psi}}' - \boldsymbol{\psi}', \widetilde{\upsilon} - \upsilon)'$ in the limit as $T \to \infty$ being multivariate normal with zero mean vector and covariance matrix equal to the inverse of (9). The necessary conditions are satisfied primarily because derivatives of the elements of $\boldsymbol{\psi}$, which are associated with the mean, depend only on sines and cosines and so are bounded; see Harvey (2013), pp 40–44. Note that condition (v) in Jensen and Rahbek (2004), p 1206, requires boundedness of the third derivatives of the log-likelihood. Derivatives with respect to concentration are also bounded because

$$\frac{\partial A(\upsilon)}{\partial \upsilon} = 1 - A(\upsilon)^2 - \frac{A(\upsilon)}{\upsilon}, \quad \upsilon > 0.$$

and cross-derivatives with respect to υ and elements of $\boldsymbol{\psi}$ are all zero.

Unfortunately the initial value of the random variable, $\mu_{1|0}$, is not usually known and the filter used in practice is an equation of the form (7), but with $\mu_{1|0}$ set to zero. In order to establish the asymptotic properties of an ML estimator based on this filter, it must be shown to be invertible. A sufficient condition for invertibility, given by (iv) of Proposition 3.2 in Blasques et al. (2022), is $\sup |z_t| < 1$, with the supremum over all admissible $\mu_{t|t-1}$, y_t and $\boldsymbol{\psi}$. This implies $|\phi - \kappa| < 1$ and $|\phi + \kappa| < 1$ because $u_t' = -\cos(y_t - \mu - \mu_{t|t-1})$ lies in the range $[-1, 1]$, with $u_t' = -1$ when $y_t = \mu + \mu_{t|t-1}$ and $u_t' = 1$ when $y_t \pm \pi = \mu + \mu_{t|t-1}$; compare the Student-$t$ location example in Blasques et al. (2022), p 337. The condition $|\phi - \kappa| < 1$ is the standard invertibility condition for a Gaussian model, (5).

Proposition 7. *When $|\phi \pm \kappa| < 1$, $\kappa \neq 0$ and $0 < \upsilon < \infty$, the model of Proposition 2 is globally identifiable and the ML estimators are consistent, with the distribution of $\sqrt{T}(\widetilde{\boldsymbol{\psi}}' - \boldsymbol{\psi}', \widetilde{\upsilon} - \upsilon)'$ in the limit as $T \to \infty$ being multivariate normal with zero mean vector and covariance matrix equal to the inverse of (9).*

Once the invertibility conditions have been established, it is not difficult to check the assumptions in Blasques et al. (2022), Section 4, for Consistency and Identifiability; see the on-line Appendix. As regards the information matrix, the invertibility conditions guarantee that $b < 1$ because if $|z_t| < 1$ for all t it follows that $z_t^2 < 1$ and so $E(z_t^2) = b < 1$. When, as will usually be the case, $\phi \geq 0$ and $\kappa > 0$, it follows that $|\phi + \kappa| \geq |\phi - \kappa|$ and so the only binding restriction is $\kappa < 1 - \phi$. This constraint may be somewhat restrictive in practice and so it is important to bear in mind that it is sufficient but not necessary. The Monte Carlo experiments reported in Table 1 are based on 10,000 replications with the first fifty observations discarded to remove the effect of initialization. The asymptotic variances are given by $T^{-1}\mathbf{I}^{-1}(\boldsymbol{\psi})$, where $\mathbf{I}(\boldsymbol{\psi})$ is in (9). It is reassuring that sample mean square errors for the ML estimates, obtained using a filter with $\mu_{1|0}$ set to zero, converge towards the asymptotic variances of the true model as T increases, even though models 2 and 3 have $\phi + \kappa > 1$. However, the optimization was started from close to the true values[2] of the parameters in $\boldsymbol{\psi}$. With real data it is prudent to check for local optima.

The ML estimates can be obtained by maximizing

$$S(\boldsymbol{\psi}) = \sum_{t=1}^{T} \cos(y_t - \mu - \mu_{t|t-1}), \quad \mu_{1|0} = 0,$$

with respect to $\boldsymbol{\psi}$ and μ. This may be done independently of υ. An initial estimate of μ is given by the sample mean direction, defined as $\bar{y}_d = \arctan^*(\bar{S}/\bar{C})$, where $\bar{S} = \sum \sin y_t/T$ and $\bar{C} = \sum \cos y_t/T$, and $\arctan^*(.)$ is as defined in Ley and Verdebout (2017), pp 10–11. Once $\widetilde{\boldsymbol{\psi}}$ has been computed, the ML estimate of υ can be obtained by solving $A(\widetilde{\upsilon}) = S(\widetilde{\boldsymbol{\psi}}, \widetilde{\mu})/T$. Unfortunately there is no exact solution; see Mardia and Jupp (2000), pp 85–86.

When a constant is added to all the observations, the likelihood is maximized simply by adding the same constant to the estimate of μ; the estimates of the other parameters remain unchanged. Thus it does not matter where we choose to locate zero on the circle when producing measurements.[3] over $[-\pi, \pi)$ or $[0, 2\pi)$. This contrasts with LARMA and IAR models where it is usually necessary to first subtract the mean direction and then wrap the observations; see the Black Mountain example[4] in Fisher and Lee (1994) and the Koeberg example in Section 2.5.

2.3. Tests for dynamic specification

The Lagrange multiplier (LM) test against serial correlation in location is based on the portmanteau or Box–Ljung statistic constructed from the autocorrelations of the scores; see Harvey (2013), pp 52–54. For a vM distribution with $\upsilon > 0$, the scores are proportional to the sines of the angular observations measured as deviations from their mean direction, so the sample autocorrelations are

$$r_c(\tau) = \frac{\sum \sin(y_t - \bar{y}_d)\sin(y_{t-\tau} - \bar{y}_d)}{\sum \sin^2(y_t - \bar{y}_d)}, \qquad \tau = 1, 2, \ldots, \tag{11}$$

[2] To be precise the true values were perturbed by random uniform variables in the range [0.95 1.05].

[3] The most convenient choice is usually the one that is directly related to the original measurements in degrees; for example $[0, 360)$ is converted to $[0, 2\pi)$ simply by multiplying by $2\pi/360$.

[4] Harvey et al. (2019) show that the score-driven model gives a better fit than the models fitted by Fisher and Lee.

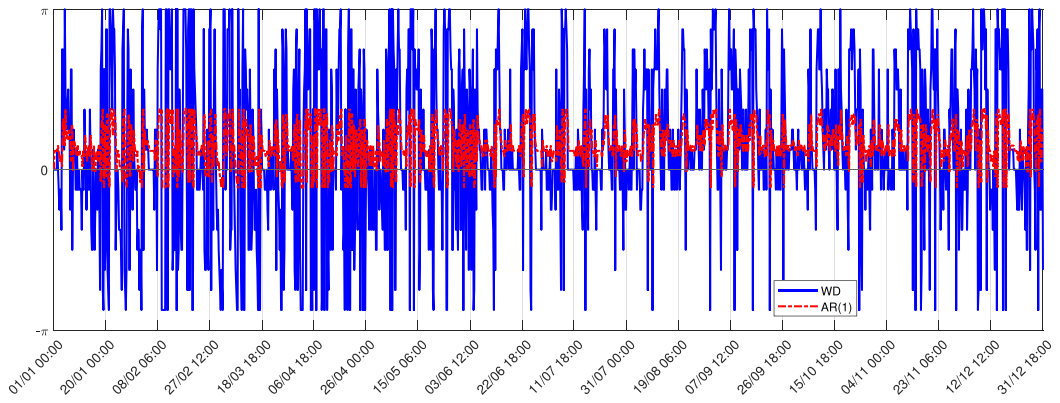

Fig. 2. Wind direction at Koeberg every 6 h, from January 1st, 1985, at 00.00 to December 31st, 1985, at 18.00. The dashed line shows the prediction from the Gaussian AR(1) model.

which correspond to the circular autocorrelations as in Jammalamadaka and SenGupta (2001), pp 176–9. The limiting distribution when the observations are independent and identically distributed (IID) is standard normal, that is $\sqrt{T}r_c(\tau) \rightarrow N(0, 1)$; see Brockwell and Davis (1991), Theorem 7.7.2 . The derivation is based on a score-driven model in which $\mu_{t|t-1}$ in (6) depends on lags of u_t. When the Q -statistic in the portmanteau test is based on the first P sample autocorrelations, it is asymptotically distributed as χ^2_P under the null hypothesis of serial independence. Once a dynamic model has been fitted, a formal test requires the degrees of freedom to be adjusted by subtracting the number of estimated dynamic parameters from P. This is the Box–Pierce test. An alternative is to carry out an LM test.

2.4. Goodness of fit and forecasts

Goodness of fit may be assessed by the dispersion (circular variance)

$$D = 1 - \sum_{t=1}^{T} \cos(y_t - \mu - \mu_{t|t-1})/T, \quad 0 \le D \le 1, \tag{12}$$

or the circular standard deviation, $s = \sqrt{-2\ln(1-D)}$, a measure whose square is most comparable to the prediction error variance; see Mardia and Jupp (2000), pp 18–19, 30. The raw dispersion, D_0, is as in (12) but with $\mu_{t|t-1}$ replaced by the mean direction, $\bar{y}_d$. Hence goodness of fit for a particular model might be characterized by a measure similar to the coefficient of multiple correlation, namely $A = 1 - D/D_0$; a perfect fit gives $A = 1$.

In time series forecasting the random walk often provides a useful benchmark. The equivalent benchmark for dispersion is $D_\Delta = 1 - \sum_{t=2}^{T} \cos(y_t - y_{t-1})/(T-1)$. Hence goodness of fit is $A_\Delta = 1 - D/D_\Delta$; a negative value indicates that the model forecasts are worse than those given by the last observation.

The same measures may be used to assess the accuracy of predictions with $D(\ell) = 1 - \sum_{j=1}^{\ell} \cos(y_{T+j} - \mu_{T+j|T})/\ell$, $D_0(\ell) = 1 - \sum_{j=1}^{\ell} \cos(y_{T+j} - \bar{y}_d)/\ell$ and $D_\Delta(\ell) = 1 - \sum_{j=1}^{\ell} \cos(y_{T+j} - y_T)/\ell$.

2.5. Example

Zucchini et al. (2017) give hourly values of the wind direction at the Koeberg nuclear power station in South Africa, classified into the 16 conventional directions N, NNE,…, NNW, and coded 1–16, in that order. The period covered is 1 May 1985 to 30 April 1989. Holzmann et al. (2006) and Zucchini et al. (2017) classified the data into 16 groups because they wanted to fit a Markov switching model. Switching has considerable merit in some circumstances and we will investigate score-driven switching models in a later paper.

The aim here is not to build a comprehensive model for hourly observations, which might include diurnal and seasonal effects,[5] but rather to use the data to illustrate the way in which the basic score-driven model handles a situation where there are observations close to π and $-\pi$. To simplify matters we take observations every six hours for the first year, giving $T = 1460$. Fig. 2 shows the observations together with the one-step ahead predictions from a Gaussian AR(1) model. It is clear there is a problem because the autoregressive coefficient is only 0.26 and the predictions are never far from zero. Fitting the score-driven (DCS) model, (6) and (7), gives the estimates in the first row of Table 2. In this case $\widetilde{\phi} = 0.90$.

[5] The possibility of systematic diurnal movements arises because the prevailing wind at night may not be the same as in the day; see Yeh et al. (2013).

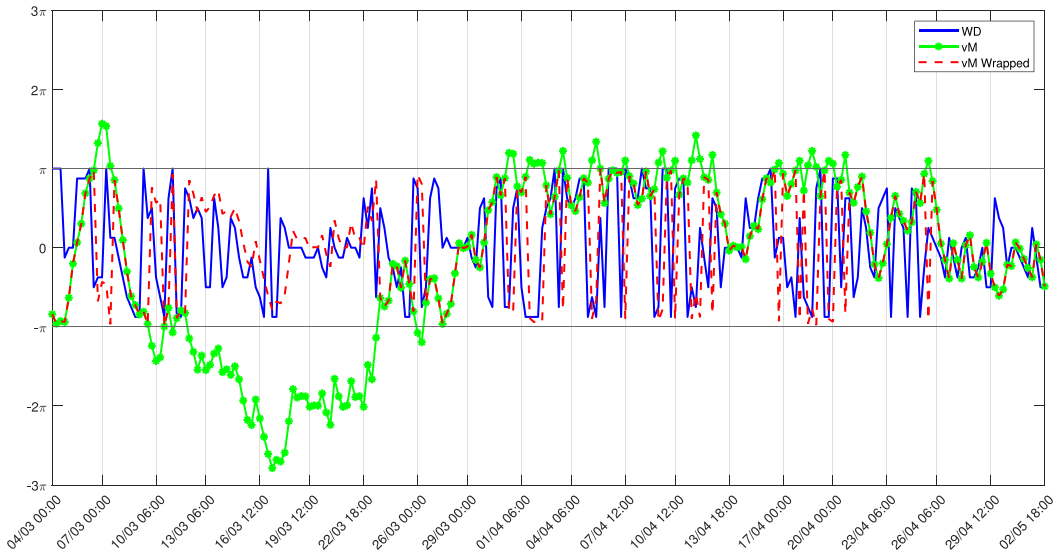

Fig. 3. Wind direction at Koeberg from March 4th, 1985, at 00.00 to May 5th at 18.00, together with DCS filter and wrapped DCS filter (dashes).

To construct a proper benchmark, the mean direction was subtracted and the observations wrapped before fitting a LARMA model using a logistic transformation. A simple LAR(1) gave as good a fit as an LARMA(1,1). The DCS model was estimated on the wrapped data to make the point that the estimates of ϕ and κ are unaffected. The estimate of the autoregressive parameter in the LAR(1) model is much smaller than it is in the score-driven model, just as it was in the naive Gaussian model. As a result the variation of the filter (one-step ahead predictions) around the mean direction is again relatively small. As can be seen from Table 2, D is smaller, and hence A bigger, for the DCS model.

Fig. 3 shows that the unwrapped DCS filter, $\mu_{t|t-1}$, frequently moves outside the range $[-\pi, \pi)$ but, because the model is stationary, it always returns. Wrapping the filter makes it follow the observations more closely, but the fit, as measured by (12), is unaffected. The graph is restricted to observations from March 4th to May 5th for clarity, but a graph for the full dataset tells the same story.

The last two columns in Table 2 compare forecasting performance over the last 240 observations based on estimates obtained without using these observations. As with the in-sample fit, the DCS model performs better than the LAR(1).

3. Non-stationary models

In a stationary score-driven model, the conditional mean can go outside the range of the observations, but it must revert towards the mean direction. When this is no longer the case, the conditional mean can, in principle, travel round and round the circle. Very little attention has been paid to modelling such situations, a recent exception[6] being Yeh et al. (2013). After setting out the theory for non-stationary score-driven models, we present two applications that illustrate how non-stationarity can be detected and handled.

3.1. Random walks and trends

A basic non-stationary model sets $\mu = 0$ in (6) and takes $\mu_{t|t-1}$ to be a random walk, that is

$$\mu_{t+1|t} = \mu_{t|t-1} + \kappa u_t, \quad t = 1, \ldots, T, \tag{13}$$

with $\mu_{1|0}$ fixed and known. Because the filter and the dynamic equation in the model are the same, and because the derivatives of $\mu_{t|t-1}$ with respect to both κ and υ are strictly stationary stochastic recurrence equations when $b < 1$, it is possible to establish the following result; compare Harvey (2013), pp 45–6.

Proposition 8. *When $\kappa > 0$, $0 < \upsilon < \infty$ and $b < 1$, the ML estimator, $(\widetilde{\kappa}, \widetilde{\upsilon})'$, is locally identifiable and consistent, with the distribution of $\sqrt{T}(\widetilde{\kappa} - \kappa, \widetilde{\upsilon} - \upsilon)'$ in the limit as $T \to \infty$ being multivariate normal with zero mean vector and covariance matrix equal to the inverse of (9) for $(\kappa, \upsilon)'$.*

[6] The model considered by Yeh et al. (2013) does not produce a likelihood function in closed form; it needs to be evaluated at each time point by numerical integration. In addition the model does not generalize easily.

The equation for b, that is (10), implies $0 < \kappa < 2A(\upsilon)\upsilon/(\upsilon - A(\upsilon))$ and so $\widetilde{\kappa}$ is asymptotically normal with mean κ and

$$\text{Avar}(\widetilde{\kappa}) = \frac{1}{T}\frac{2\kappa A(\upsilon) - \kappa^2(1 - A(\upsilon)/\upsilon)}{A(\upsilon)^2}. \tag{14}$$

Note that $2A(\upsilon)\upsilon/(\upsilon - A(\upsilon)) \to 0$ as $\upsilon \to 0$ whereas for $\upsilon \to \infty$, $0 < \kappa < 2$. The results of a Monte Carlo experiment, again based on 10,000 replications with the first fifty observations discarded, are reported in Table 3. The DCS model of (5) and (13) was estimated by ML with $\mu_{1|0} = 0$ in the filter. The MSEs for $\widetilde{\kappa}$ are close to the values given by (14). Note that rather than begin by setting[7] $\mu_{1|0} = 0$ in the filter based on (13), it may be better to start the recursion at $t = 2$ with $\mu_{2|1} = y_1$. Another option is to treat $\mu_{1|0}$ as an unknown parameter; the transformation $\mu_{1|0} = 2\arctan(\omega)$, where ω is unconstrained, could be used to ensure that $|\widetilde{\mu}_{1|0}| < \pi$.

Extending (6) to include a deterministic time trend gives

$$x_t = \mu_{t|t-1} + \mu + \beta t + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{15}$$

where $\mu_{t|t-1}$ is as in (7). Provided $\mu_{t|t-1}$ is stationary, that is $|\phi| < 1$, and $\kappa < 1 - \phi$, the model is globally identifiable as in Proposition 3. For a time trend with no serial correlation, that is $\mu_{t|t-1} = 0$ for all t, it is straightforward to show, by following the argument in Hamilton (1994), p 460, that the ML estimators of μ and β are consistent and

$$\begin{bmatrix} \sqrt{T}(\widetilde{\mu} - \mu) \\ T^{3/2}(\widetilde{\beta} - \beta) \end{bmatrix} \xrightarrow{L} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \frac{1}{\upsilon A(\upsilon)}\begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix}^{-1}\right).$$

For the full model with a time trend and stationary first-order dynamics, that is (6) and (15), the information matrix can be derived as in Harvey (2013), pp 57–8. Deterministic periodic effects, such as seasonals and intra-day patterns, can be similarly handled.

3.2. First differences

A score-driven model for a differenced variable may be set up as

$$\Delta x_t = \beta_{t|t-1} + \beta + \varepsilon_t^{\Delta}, \qquad \beta_{2|1} = 0, \quad t = 2, \ldots, T, \tag{16}$$

where ε_t^{Δ} has a circular distribution with location zero, and

$$\beta_{t+1|t} = \phi\beta_{t|t-1} + \kappa_\beta u_t^{\Delta}, \tag{17}$$

where the score, u_t^{Δ}, is in terms of Δx_t. Because of circularity, there is no need to wrap the differenced observations, Δy_t, to ensure that they lie in an interval of 2π. Thus when ε_t^{Δ} has a von Mises distribution

$$f(y_t; \mu, \upsilon) = \frac{1}{2\pi I_0(\upsilon)}\exp\{\upsilon\cos(\Delta y_t - \beta_{t|t-1} - \beta)\}, \qquad t = 2, \ldots, T, \tag{18}$$

and

$$u_t^{\Delta} = \upsilon\sin(\Delta y_t - \beta_{t|t-1} - \beta).$$

Predictions for x_t may be obtained from (16) by the recursion

$$\widehat{x}_{T+j|T} = \widehat{x}_{T+j-1|T} + \beta_{T+j|T} + \beta, \qquad j = 1, \ldots, \tag{19}$$

where $\widehat{x}_{T|T} = y_T$. These values may need to be wrapped if they are to be compared with the original observations in the range $[-\pi, \pi)$.

3.3. Applications

In a linear Gaussian framework, the differenced model is nested within the levels model. This is no longer the case for our circular model. A viable strategy is to fit (6) with the first-order component, (7). If the estimate of ϕ is close to one, the plot of $\mu_{t|t-1}$ helps the decision as to whether or not to fit the first difference model, (16) and (17). Supplementary evidence is provided by a 't-test' on the time trend in (15).

[7] Although this is the unconditional expectation of $\mu_{1|0}$, it is somewhat arbitrary because the model could be defined over a different 2π range.

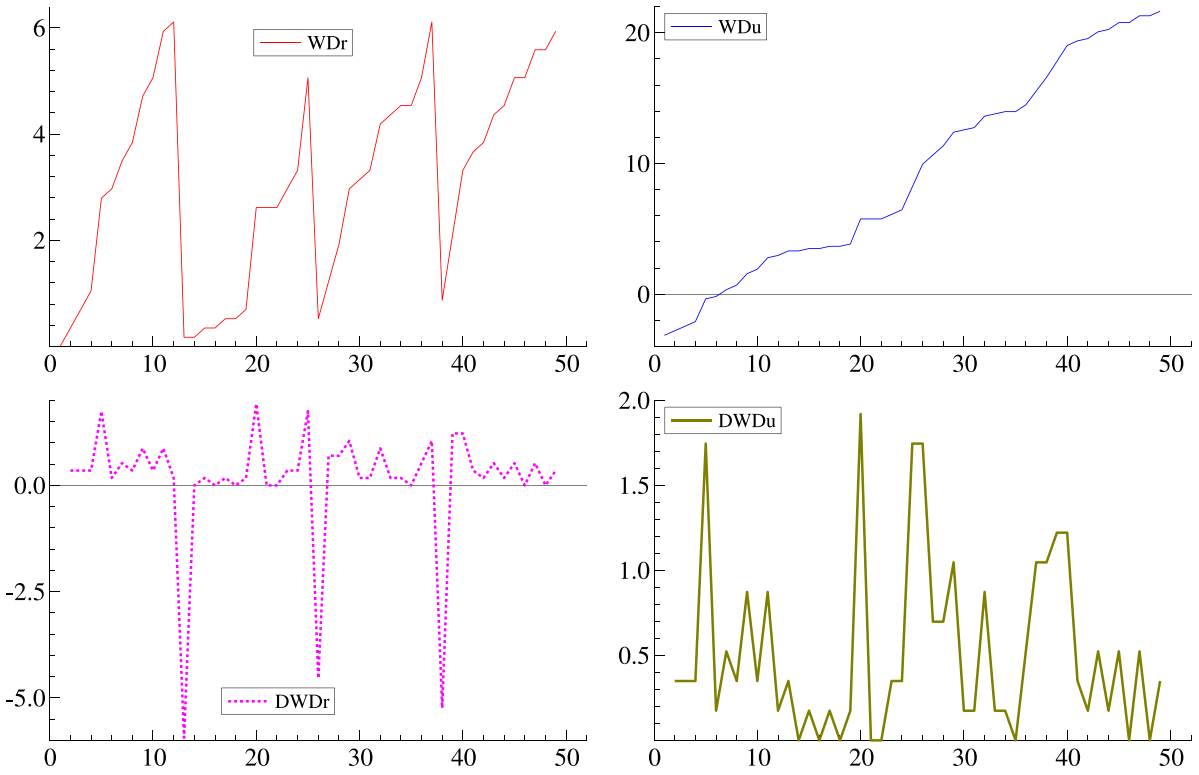

Fig. 4. Top left — weekly wind direction in Gorleston, *WDr*, in radians from zero to 2π. Top right — data with 2π added after sharp drops. Bottom left — top left differenced. Bottom right — top right differenced.

3.3.1. Gorleston

The top left hand graph of Fig. 4 shows weekly observations at Gorleston, a small coastal town in England; the source is Fisher (1993), p 249. It is immediately apparent that the wind goes all the way round the circle because there are sharp drops as the upper limit of 2π is approached. Adding 2π to all the observations after such drops gives the series in the top right hand graph. This series is clearly nonstationary and taking first differences gives the series immediately below it. Differencing the original data gives the observations in the bottom left hand graph; it is necessary to add 2π at the points where the sharp drops occur, that is wrap them, in order to get the differenced series in the right hand graph. However, as noted earlier, it makes no difference whether the score-driven model is estimated with the raw first differences in the bottom left hand graph of Fig. 4 or with the modified data to the right.

As can be seen from Table 4 the estimation of (6), with the original data, results in $\widetilde{\phi} = 1$. When a time trend is included, as in (15), $\widetilde{\phi} = 0.862$ (0.081), $\widetilde{\kappa} = 1.177$ (0.143) and $\widetilde{\beta} = 0.522$ (0.022). There is very little residual serial correlation, with a plot (not shown here) of the filtered level, $\mu + \beta t + \mu_{t|t-1}$, $t = 1, \ldots, T$, following a similar path to that of the unwrapped observations in Fig. 4. Fitting a LARMA model with a time trend to wrapped data clearly makes no sense.

When the model is estimated in first differences, with a mean and first-order dynamics, the log-likelihood increases. The serial correlation is moderate with $\widetilde{\phi} = 0.229$ (0.880) and because $\widetilde{\kappa} = 0.125$ (0.219) the invertibility condition is satisfied.

3.3.2. Ancona

The lower part of Fig. 5 shows a time series of wind direction recorded every half hour in the period from 00.00 on 17/2/2010 to 23.30 on 17/3/2010 by the buoy of Ancona, located in the Adriatic; see Lagona (2019). Fitting an IAR or LARMA model would be problematic because the observations are not confined to a narrow arc. Indeed there is some indication that the observations are moving round the circle, but making the adjustments by eye, as was done for Gorleston, is not an option because there is more variability and the series is much longer. As can be seen from Table 5 fitting[8] the score-driven model, (6), gives $\widetilde{\phi} = 0.999(0.000)$, $\widetilde{\kappa} = 1.687$ (0.018) and $\widetilde{\upsilon} = 41.45$ (0.039) and the pattern of the filtered level in the upper part of Fig. 5 shows that the wind direction is continually moving round the circle. When

[8] There are a few missing observations which are handled by setting the score to zero in the filter.

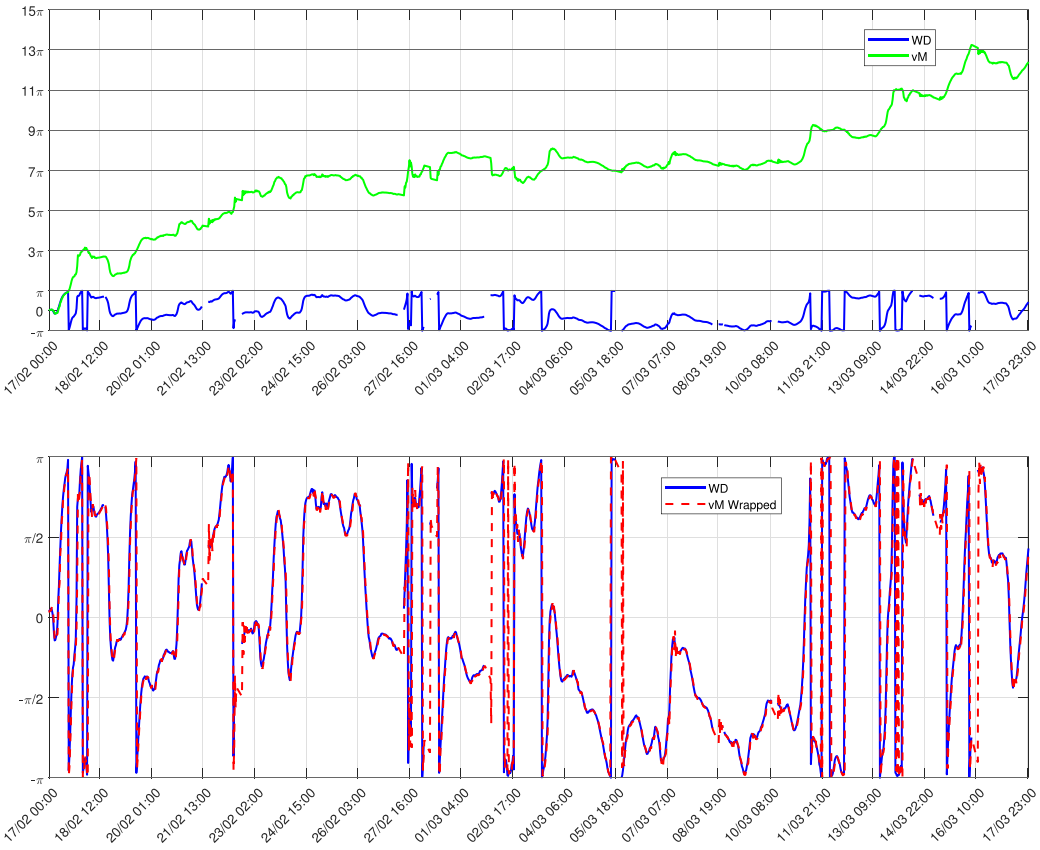

Fig. 5. Upper panel: Wind direction, in radians, recorded every half hour in the period from 0.00 am on 17/2/2010 to 23.30 on 17/3/2010 by the buoy of Ancona, together with filter. Lower panel: Wind direction and wrapped filter (dashed).

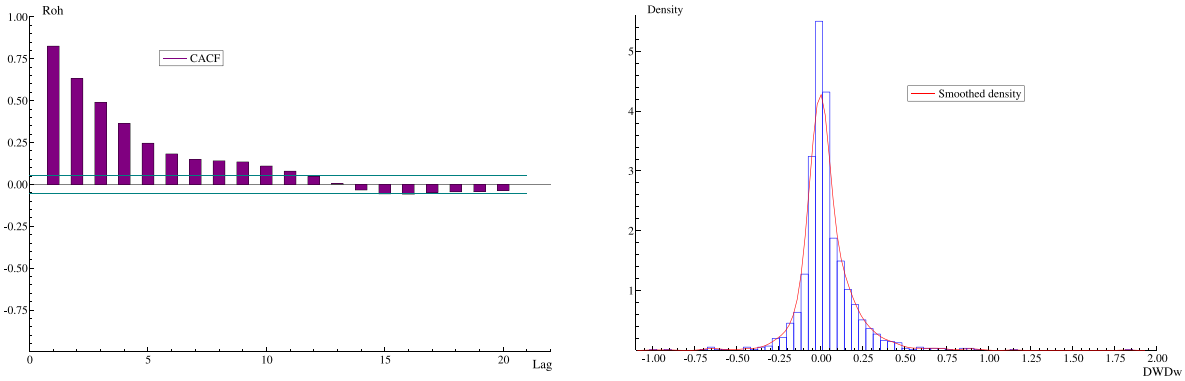

Fig. 6. Left panel: CACF of differenced observations on Ancona wind direction. Right panel: Histogram of (wrapped) differenced Ancona observations.

wrapped, the filter is close to the observations. The estimate of ϕ for the LAR(1) model is 0.810 which is, as expected, much smaller than the near unit root estimate obtained for the score-driven model.

When a time trend is included, as in (15), the results are $\widetilde{\phi} = 0.994(0.001)$, $\widetilde{\kappa} = 1.719(0.018)$, $\widetilde{\beta} = 0.021\ (0.001)$ and $\widetilde{\upsilon} = 40.53$. Although the coefficient of β is quite small, the $t-statistic$ takes the value 20.60 and so is highly significant. (When the time trend model was fitted to Koeberg, the estimate of β was -0.0001 with a $t-statistic$ of only -1.173.)

The sample circular autocorrelation function (CACF) of the differenced observations, shown in Fig. 6, exhibits strong positive serial correlation and confirms the need to model the differenced observations. The histogram of the (wrapped) differenced observations, Fig. 6, points to the presence of outliers and the possibility of skewness. The next section extends the basic vM model to accommodate these characteristics.

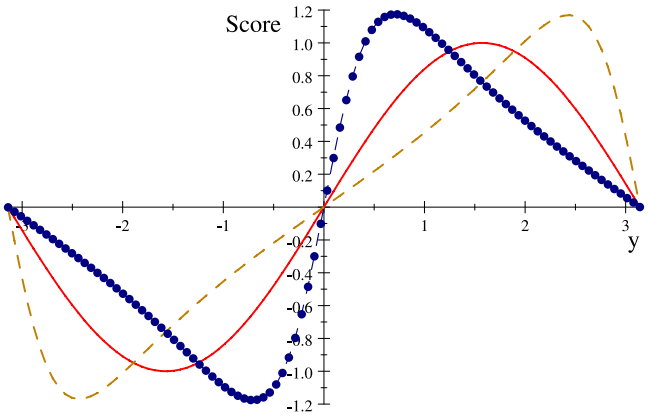

Fig. 7. Sine, Cardioid (- -) and wrapped Cauchy (dots) scores for location with $\upsilon = 1$.

4. A general class of distributions

Jones and Pewsey (2005) give a distribution that includes vM, Cardioid and wrapped Cauchy as special cases. The PDF is

$$f(y_t) \propto [1 + \tanh(\varkappa\upsilon)\cos(y_t - \mu)]^{1/\varkappa}, \quad -\pi \leq y_t, \mu < \pi, \tag{20}$$

where $\upsilon \geq 0$, $-\infty \leq \varkappa \leq \infty$. The value of $\varkappa$ is 1 for the Cardioid distribution, -1 for wrapped Cauchy (WC) and 0 for von Mises. The Cardioid and WC distributions depend on a scale parameter that can be related to υ. For the Cardioid the parameter is $\rho = \tanh(\upsilon)/2$, $0 \leq \rho \leq 1/2$, whereas for WC it is $\rho = \tanh(\upsilon/2)$, $0 \leq \rho < 1$. In a model related to the IAR(1), Kato (2010) argues the case for a conditional WC distribution in connection with a wind dataset where there are a small number of big changes in direction.

Fig. 7 contrasts the location score of a von Mises distribution, where $\upsilon = 1$, with the corresponding scores of a Cardioid and WC which are

$$u_t = \frac{2\rho\sin(y_t - \mu_{t|t-1})}{1 + 2\rho\cos(y_t - \mu_{t|t-1})} \quad \text{and} \quad u_t = \frac{2\rho\sin(y_t - \mu_{t|t-1})}{1 + \rho^2 - 2\rho\cos(y_t - \mu_{t|t-1})} \tag{21}$$

respectively. All three are unchanged when a multiple of 2π is added to, or subtracted from, the angle. The WC score reaches a turning point for $|y| < \pi/2$ and does so more rapidly as υ increases. The Cardioid score, on the other hand, reaches its turning point when $|y| > \pi/2$. The scores become similar as $\upsilon \to 0$.

Skewness may be introduced by multiplying the PDF by $1 + \varsigma\sin(y_t - \mu_{t|t-1})$, where ς is a skewing parameter, as in Abe and Pewsey (2011). Abe et al. (2017) fit such a distribution to a series on wind direction where it seems that there is a gradual rotation in an anti-clockwise direction. Skewness subtracts

$$\frac{\varsigma\cos(y_t - \mu_{t|t-1})}{1 + \varsigma\sin(y_t - \mu_{t|t-1})},$$

from the score.

Evaluation of the likelihood function for the general Jones–Pewsey distribution requires numerical integration to find the normalizing constant, which is a function of $\varkappa$ and υ. But the Cardioid and WC distributions, and the other special cases listed in Jones and Pewsey (2005), p 1423, have a simple closed form.

Table 6 shows the results of fitting vM, WC and Cardioid models to the (unwrapped) differenced observations. As might have been anticipated, the skewed WC gives the best fit, while the Cardioid gives the worst. A plot of the filtered AR component for the vM shows that it tends to follow the outliers, whereas the corresponding plot for (skewed) WC shows a much narrower range of movement that is not unduly influenced by the outliers. This is consistent with a scenario in which the wind suddenly changes direction.

5. Conclusions

This article has shown how the score-driven approach provides a natural solution to the difficulties posed by circular data and leads to a coherent and unified methodology for estimation, model selection and testing. An asymptotic theory is developed and Monte Carlo experiments show the results to be consistent with small sample performance, even when invertibility cannot be established. Diagnostic checks for serial correlation follow straightforwardly. It is argued that existing models for circular data are unable to deal with observations that do not lie on a narrow arc of the circle and an empirical example illustrates the point.

Table 1
MSEx10 for ML estimates of parameters of a vM DCS model. Asymptotic variances are shown in brackets.

ϕ	κ	υ	T	μ	ϕ	κ	υ
0.7	0.2	2	250	0.299	0.805	0.069	0.218
			500	0.030 (0.030)	0.366 (0.192)	0.037 (0.034)	0.115 (0.122)
			1000	0.014	0.166	0.017	0.059
			2000	0.007 (0.008)	0.058 (0.048)	0.009 (0.009)	0.033 (0.030)
0.7	0.5	2	250	0.127	0.132	0.077	0.260
			500	0.064 (0.064)	0.056 (0.045)	0.038 (0.037)	0.128 (0.122)
			1000	0.032	0.024	0.018	0.062
			2000	0.016 (0.016)	0.012 (0.011)	0.009 (0.009)	0.030 (0.030)
0.9	0.5	4	250	0.603	0.026	0.045	1.137
			500	0.204 (0.162)	0.009 (0.007)	0.022 (0.022)	0.541 (0.520)
			1000	0.084	0.004	0.011	0.265
			2000	0.041 (0.040)	0.002 (0.002)	0.005 (0.005)	0.131 (0.130)

Table 2
Results of fitting the following models to the Koeberg dataset: von Mises DCS, labelled (1) and (3), Gaussian AR(1), labelled (2), and LAR(1), labelled (4). In fitting (3) and (4) we have first subtracted the mean direction and then wrapped the data. Standard errors are in brackets.

Model	μ	ϕ	κ	υ	σ	Fit		Forecast	
						D	A	D	A
(1)	0.5274 (0.0673)	0.9263 (0.0022)	1.2755 (0.0138)	1.1431 (0.0403)		0.5052	0.3427	0.5429	0.3107
(2)	0.3685 (0.0018)	0.2616 (0.0006)			1.5304	0.6717	0.1261	0.6754	0.1424
(3)	−0.1102 (0.0673)	0.9263 (0.0022)	1.2755 (0.0138)	1.1431 (0.0403)		0.5052	0.3427	0.5359	0.3196
(4)	0.1061 (0.0011)	0.3485 (0.0006)			1.2835	0.5545	0.2785	0.5531	0.2978

The variable in question can sometimes continue to move round the circle. Score-driven models can deal with such situations and in doing so they are able to reveal the nonstationarity in the data and inform the decision as to whether or not to take first differences. The applications presented in the paper show our proposed strategy works in practice.

Finally the score-driven model developed for a von Mises distribution is extended to the more general Jones–Pewsey class. Particular attention is paid to the wrapped Cauchy distribution, which, as we show, is able to fit datasets where there are sudden changes in direction.

The models can be further developed to accommodate heteroscedasticity, with dynamic equations driven by the score with respect to concentration, and to include seasonal and diurnal effects. They also offer the possibility of handling switching regimes and constructing dynamic models of speed and direction based on distributions for a cylinder, such as the one proposed by Abe and Ley (2017). These issues will be addressed in later work.

Acknowledgements

We are grateful to Francisco Lagona for supplying the Ancona data. Earlier versions of some of the ideas in this paper were presented at the INET workshop on Score-driven and Nonlinear Time Series Models held at Trinity College, Cambridge, in March 2019, at the Econometric Models of Climate Change conferences in Milan in August 2019 and on-line in August 2021, at the 22nd Oxmetrics conference at Nuffield college, Oxford in September, 2019, and at workshops in Cambridge, Bologna, Ecole Polytechnique Féd érale de Lausanne, the QUT Centre for Data Sciences and the University of Konstanz. We wish to thank Richard Davis, Anthony Davison, Jurgen Doornik, David Hendry, Peter Jupp, Rutger-Jan Lange, Ken Lindsay, Oliver Linton, Alessandra Luati, Alexei Onatski, Neil Shephard, Richard Smith, Howell Tong and others for helpful comments. We would also like to thank the referees for some important and insightful comments, as well as suggestions for clarifying some of the results. Any errors are our responsibility. This work was developed within the project funded by Next Generation EU - Project PE0000018 "GRINS - Growing Resilient, INclusive and Sustainable", National Recovery and Resilience Plan (NRRP) - EP9 - Mission 4, C2, Intervention 1.3. The views and opinions expressed are only those of the authors and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

Appendix A. Tables

See Tables 1–6.

Table 3
MSEx10 for ML estimates of parameters of a vM DCS model with conditional mean given by a random walk. Asymptotic variances in brackets.

T	$\kappa = 1$	$\upsilon = 2$
250	0.066 (0.061)	0.255 (0.244)
500	0.032 (0.031)	0.127 (0.122)
1000	0.017 (0.015)	0.062 (0.062)
2000	0.009 (0.008)	0.030 (0.030)

Table 4
Results of fitting the following models to the Gorleston dataset: DCS, labelled (1); DCS as in (1) but with time trend, (2); DCS on first differences, (3). BIC is Bayes information criterion and lnL is maximized log-likelihood. Standard errors are in brackets.

Model	μ	ϕ	κ	β	υ	BIC	lnL	D	A	A_Δ
y_t										
(1)	−2.9812 (0.5748)	1.0000 (0.0388)	1.7525 (0.4288)		3.0415 (0.4979)	110.04	- 47.24	0.1834	0.7915	0.1739
(2)	−3.1416 (0.4568)	0.8620 (0.0813)	1.1767 (0.1430)	0.5215 (0.0215)	4.9198 (0.1904)	85.22	- 32.88	0.1063	0.8792	0.5214
Δy_t										
(3)	0.5835 (0.0885)	0.2289 (0.8796)	0.1247 (0.2190)		4.6880 (0.2061)	72.55	- 28.53	0.1148	0.0399	0.3862

Table 5
Results of fitting the following models to the Ancona dataset: DCS, labelled (1); DCS as in (1) but with time trend, (2); Gaussian AR(1), labelled (3), and LAR(1), labelled (4). Standard errors are in brackets.

Model	μ	ϕ	κ	β	υ	σ	D	A	A_Δ
y_t									
(1)	0.0955 (0.1107)	0.9987 (0.0001)	1.7242 (0.0184)		38.8768 (0.0391)		0.0824	0.9261	0.1189
(2)	0.1356 (0.1091)	0.9938 (0.0008)	1.7191 (0.0179)	0.0210 (0.0010)	40.5259 (0.0391)		0.0819	0.9265	0.1241
(3)	−0.0065 (0.0005)	0.9105 (0.0001)				0.7953	0.1009	0.9096	−0.0780
(4)	−0.0239 (0.0010)	0.8102 (0.0002)				1.1779	0.1128	0.8989	−0.2054

Table 6
First-order DCS models fitted to the differenced Ancona observations under various distributional assumptions. A and D are now for the differenced observations. Standard errors are in brackets.

Model	μ	ϕ	κ	υ/ρ	ζ	BIC	lnL	D	A
Δy_t									
Static Von Mises	0.0325 (0.0047)			36.2051 (0.0396)		- 912.16	463.32	0.0931	0.0000
Von Mises	0.0307 (0.0115)	0.7287 (0.0244)	0.9309 (0.0323)	96.7895 (0.0398)		-2,145.00	1,087.00	0.0852	0.0845
Skewed Von Mises	0.0114 (0.0269)	0.7276 (0.0245)	0.9308 (0.0322)	96.8377 (0.0398)	0.1826 (0.0180)	-2,138.37	1,087.28	0.0852	0.0845
Wrapped Cauchy	−0.1737 (0.0025)	0.9811 (0.0004)	0.0035 (0.0000)	0.9465 (0.0024)		-2,523.20	1,276.10	0.0880	0.0553
Skewed Wrapped Cauchy	0.0269 (0.0003)	0.9201 (0.0000)	0.0033 (0.0000)	0.9496 (0.0001)	1.0000 (0.6117)	-2,621.20	1,328.70	0.0879	0.0562
Cardioid	0.0346 (0.1485)	0.7445 (0.3274)	1.5525 (0.6750)	0.5000 (9.8144)		2,914.40	-1,442.70	0.0856	0.0806
Skewed Cardioid	−0.7490 (0.0936)	0.7528 (0.2050)	0.6086 (0.1508)	0.5000 (7.6577)	1.0000 (6.6280)	1,986.40	- 975.12	0.3532	−2.7935

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2023.02.016.

References

Abe, T., Ley, C., 2017. A tractable, parsimonious and flexible model for cylindrical data, with applications. Econom. Stat. 4, 91–104.
Abe, T., Ogata, H., Shiohama, T., Taniai, H., 2017. Circular autocorrelation of stationary circular Markov processes. Stat. Inference Stoch. Process. (ISSN: 1387-0874) 20 (3), 275–290.

Abe, T., Pewsey, A., 2011. Sine-skewed circular distributions. Statist. Papers 52 (3), 683–707.
Blasques, F., van Brummelen, J., Koopman, S.J., Lucas, A., 2022. Maximum likelihood estimation for score-driven models. J. Econometrics 227 (2), 325–346.
Breckling, J., 1989. The Analysis of Directional Time Series: Applications to Wind Speed and Direction. Springer New York, NY.
Brockwell, P., Davis, R., 1991. Time Series: Theory and Methods. Springer New York, NY.
Clarke, H., Lucas, C., Smith, P., 2013. Changes in Australian fire weather between 1973 and 2010. Int. J. Climatol. 33 (4), 931–944.
Coles, S., 1998. Inference for circular distributions and processes. Stat. Comput. 8, 105–113.
Creal, D., Koopman, S.J., Lucas, A., 2013. Generalized autoregressive score models with applications. J. Appl. Econometrics 28, 777–795.
Erdem, E., Shi, J., 2011. ARMA based approaches for forecasting the tuple of wind speed and direction. Appl. Energy 88 (4), 1405–1414.
Fisher, N.I., 1993. Statistical Analysis of Circular Data. Cambridge University Press.
Fisher, N., Lee, A., 1994. Time series analysis of circular data. J. R. Stat. Soc. Ser. B Stat. Methodol. 56 (2), 327–339.
García-Portugués, E., Crujeiras, R., González-Manteiga, W., 2013. Exploring wind direction and SO2 concentration by circular linear density estimation.. Stoch. Environ. Res. Risk Assess. (27), 1055–1067.
Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, New Jersey.
Harvey, A.C., 2013. Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series. Cambridge University Press.
Harvey, A., Hurn, S., Thiele, S., 2019. Modeling Directional (Circular) Time Series. Cambridge Working Papers in Economics (1971).
Holzmann, H., Munk, A., Suster, M., Zucchini, W., 2006. Hidden Markov models for circular and linear-circular time series. Environ. Ecol. Stat. 13, 325–347.
Hurn, S., Tian, J., Xu, L., 2021. Assessing the informational content of official Australian bureau of meteorology forecasts of wind speed. Econ. Rec. 97 (319), 525–547.
Jammalamadaka, S., SenGupta, A., 2001. Topics in Circular Statistics. World Scientific Publishing Co., New Jersey.
Jensen, S.T., Rahbek, A., 2004. Asymptotic inference for nonstationary GARCH. Econom. Theory 20 (6), 1203–1226.
Jones, M., Pewsey, A., 2005. A family of symmetric distributions on the circle. J. Amer. Statist. Assoc. (100), 1422–1428.
Kato, S., 2010. A Markov process for circular data. J. R. Stat. Soc. Ser. B Stat. Methodol. 72 (5), 655–672.
Lagona, F., 2019. Copula-based segmentation of cylindrical time series. Statist. Probab. Lett. (144), 16–22.
Ley, C., Verdebout, T., 2017. Modern Directional Statistics. CRC Press, New York.
Lobeto, H., Menendez, M., Losada, I., Hemer, M., 2022. The effect of climate change on wind-wave directional spectra. Glob. Planet. Change 213, 103820.
Mardia, K.V., Jupp, P., 2000. Directional Statistics. Wiley, New Jersey.
Page, W.G., Wagenbrenner, N.S., Butler, B.W., Forthofer, J.M., Gibson, C., 2018. An evaluation of NDFD weather forecasts for wildland fire behavior prediction. Weather Forecast. 33 (1), 301–315.
Pewsey, A., García-Portugués, E., 2021. Recent advances in directional statistics. TEST Off. J. Span. Soc. Stat. Oper. Res. 30 (1), 1–58.
Rapp, C.E., Wilson, R.S., Toman, E.L., Jolly, W.M., 2021. Assessing the role of short-term weather forecasts in fire manager tactical decision-making: a choice experiment. Fire Ecol. 17 (1), 1–17.
Sharples, J.J., McRae, R.H., Weber, R., 2010. Wind characteristics over complex terrain with implications for bushfire risk management. Environ. Model. Softw. 25 (10), 1099–1120.
Song, D., Yang, J., Liu, Y., Su, M., Liu, A., Joo, Y.H., 2017. Wind direction prediction for yaw control of wind turbines. Int. J. Control Autom. Syst. 15 (4), 1720–1728.
Straumann, D., Mikosch, T., 2006. Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. Ann. Statist. 34 (5), 2449–2495.
van Oldenborgh, G.J., Krikken, F., Lewis, S., Leach, N.J., Lehner, F., Saunders, K.R., van Weele, M., Haustein, K., Li, S., Wallom, D., Sparrow, S., Arrighi, J., Singh, R.K., van Aalst, M.K., Philip, S.Y., Vautard, R., Otto, F.E.L., 2021. Attribution of the Australian bushfire risk to anthropogenic climate change. Nat. Hazards Earth Syst. Sci. 21 (3), 941–960.
Yeh, S., Harris, K., Jupp, P., 2013. A drifting Markov process on the circle, with physical applications. Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. 469 (2156), 20130092.
Zucchini, W., MacDonald, I.L., Langrock, R., 2017. Hidden Markov Models for Time Series: An Introduction using R. Monographs on Statistics and Applied Probability, CRC press, New York.